A STATISTICAL ASSESSMENT TOOL FOR ELECTRICITY DISTRIBUTION NETWORKS

Sathsara Abeysinghe

Institute of Energy, School of Engineering

Cardiff University

A thesis submitted for the degree of

Doctor of Philosophy

March 31, 2018

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support of many individuals. First and foremost, I offer my sincerest gratitude to my supervisor, Professor Jianzhong Wu, for his tremendous support, guidance and patience throughout my research study. I deeply appreciate the moral support and the freedom he gave me, to move on during the most difficult times.

I gratefully acknowledge the funding received towards my PhD and the internship opportunity I received from the Toshiba TRL, Bristol, UK. Many thanks to my industrial supervisor, Professor Mahesh Sooriyabandara for all his support and encouragement.

My deepest appreciation goes to Professor Nick Jenkins and Professor Janaka Ekanayake, whose advice and scientific guidance helped me greatly in conducting my research. I would also like to thank Dr. Meysam and Dr. Carlos for their input and feedback for my research.

A special thank you to Dr. Xu Tao, Professor Chengshan Wang, Professor Hongjie Jia and to all other Chinese friends I met in Tianjin, for making those five months of data collection in China all the more interesting.

I would like to thank, Dr. Orestis Georgiu and Dr. Aftab Khan for their support during the internship time in Toshiba TRL.

I'm also thankful to Miss. Qi Qi for her hard work and patience during our collaboration work.

I will always be grateful to Dr. Muditha Abeysekera for his support in proofreading my thesis and also for suggesting many improvements to the content.

To my friend Rui Dantas, thank you for listening, offering me advice, and supporting me through this entire process. I thank all my friends and colleagues who helped in many ways during these four years in Cardiff.

A heartfelt thank you to my Mother, Father and Sisters for always believing in me and encouraging me to follow my dreams. Finally, a special thank you to Krzysztof, who has been by my side believing in me and always showing how proud he is of me. Your support during the final stages of this PhD. helped me greatly. Thank you.

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed …… (candidate) Date ……

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed …… (candidate) Date ……

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third Party Editors by Research Degree Students. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed …… (candidate) Date ……

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signed …… (candidate) Date ……

ABSTRACT

With a large penetration of low carbon technologies (LCTs) (E.g. solar photovoltaics, wind turbines and electric vehicles) at medium voltage and low voltage levels, electricity distribution networks (EDNs) are undergoing rapid changes. Research has been carried out to analyse and quantify the impacts of LCTs on EDNs. Most of these previous studies are based on either real or synthetic network samples. The results and conclusions derived from these studies have limited applicability to other networks, thus making it difficult to arrive at generalised and robust conclusions on the impact of LCTs on EDNs. One of the main reasons for using a case study or synthetic networks in research is limited accessibility to real-world network data.

To bridge this research gap, the rationale and the development of a network modelling tool that can generate random-realistic representations of different types (sub-urban/urban) of EDNs in aiding statistical analysis of the power networks is presented in this thesis. The ability to generate ensembles of statistically-similar distribution networks is one of the key properties of the proposed tool. Statistically-similar distribution networks are a set networks with a similar set of topological and electrical properties as defined by the user with some given values or ranges of values. As part of this thesis four key contributions are presented.

(i) An investigation of the topological properties of real-world EDNs: The key topological properties that characterise different types of EDNs were identified and quantified. A novel depth dependent approach was developed to investigate the network topologies.
(ii) An investigation of electrical properties of real-world EDNs: The key electrical properties that characterise different types of EDNs were identified and quantified. A novel depth dependent approach was developed to investigate the electrical properties of EDNs.
(iii) Development of a statistically-similar networks generator (SSNG): A SSNG was developed as a data driven model. Real-world network properties as characterised by the previous topological and electrical investigations and the corresponding network planning and design guidelines were used in the development of the SSNG.
(iv) The application of the SSNG to analyse the impacts of soft open points (SOPs) on EDNs: A statistical analysis of the impact of Soft Open Points (SOPs) on a set of statistical-similar EDNs with variable distributed generation penetration was presented.

The developed SSNG has been validated through a statistical analysis of the impact of SOPs, and the results showed that the SSNG is able to provide robust and generalised conclusions on distribution network studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

Abbreviations

BFS	Breadth First Search
CDF	Cumulative Distribution Function
CNA	Complex Network Analysis
DER	Distributed Energy Resources
DG	Distributed Generation
DNO	Distribution Network Operator
EDN	Electricity Distribution Networks
EV	Electric Vehicles
GB	Great Britain
HV	High Voltage
LCT	Low Carbon Technologies
LV	Low Voltage
MV	Medium Voltage
NOP	Normally Open Point
OH	Over Head Lines
PDF	Probability Density Function
PMF	Probability Mass Function
PV	Photovoltaics
PS	Primary Substation
PSO	Particle Swarm Optimization
SOP	Soft Open Point
SS	Secondary Substation
SSNG	Statistically-Similar Networks Generator
UG	Under Ground Cables
VSC	Voltage Source Converter

Sets and matrices

$\boldsymbol{A}$	Adjacency matrix
$\boldsymbol{B}$	Branch matrix
$\boldsymbol{D}$	Distance matrix
$\boldsymbol{E}$	Edges set
$\boldsymbol{ED}$	Electrical distance matrix
$\boldsymbol{G}$	Graph
$\boldsymbol{K}$	Degree matrix
$\boldsymbol{L}$	Laplacian matrix
$\boldsymbol{PS}$	Primary substations list
$\boldsymbol{SS}$	Secondary substations matrix
$\boldsymbol{V}$	Vertices/nodes set

Parameters and variables in Chapter 3

b_r	Branching rate
C	Average clustering coefficient
C_i	Clustering coefficient of node i
$circuit\_id$	Circuit identification number
D_f	Fractal dimension
d_{max}	Maximum depth of the network
$d_{r,x}$	Depth of node x from the root node r
e_{avg}	Average edge length (km)
end_x, end_y	Coordinates of an end point of an edge
k_{avg}	Average degree of the network
k_i	Degree of a node i
l	Level of the network
l_{avg}	Average path length (km)
L_{total}	Total network length (km)
M	Total number of edges

m_e	Number of edges in the edge length range e
N	Total number of nodes
n_k	Number of nodes with degree equals to k
$P(e)$	Edge length distribution
$P(k)$	Degree distribution
r	Root node
$start_x, stary_y$	Coordinates of a starting point of an edge
x, y	Coordinates of the nodes
ε	Box size used in the box counting method
$\lambda_G(i)$	Number of edges between the neighbours of node i
ρ	Pearson correlation coefficient
$\tau_G(i)$	Total number of edges that can exist among the neighbours of node i

Parameters and variables in Chapter 4

$d(i,j)$	Length of the distribution line connecting nodes i and j (km)
$dist_{s,SS}$	Distance from source node to the secondary substation along the feeder (km)
ed_{avg}	Average electrical distance between two nodes of the network (Ohms)
$f_h(y)$	Kernel density probability distribution function
h	Bandwidth used in the kernel density function
k_d	Demand factor of a secondary substation
$K(y)$	Kernel smoothing function
LB_{branch}	Branch load balancing index
LB_{sys}	Load balancing index of the whole network
LD	Load density (MVA/km^2)
$level$	Level of the network
l_{max}	Maximum feeder length of the network (km)

l_{norm}	Normalised distance from source node to secondary substations
M	Total number of branches/edges in the network
m	Number of levels in a network
N	Total number of nodes in the network
n	Total number of secondary substations in a level
$node\_id$	Node identifiers
N_{SS}	Total number of secondary substations in the network
$num\_of\_feeders$	Number of outgoing feeders from the primary substation
$Power\ loss\ ratio$	The ratio of total power loss to the total supplied power of a network
PS_c	Installed capacity of a primary substation (MVA)
S_{ij}, P_{ij}, Q_{ij}	Apparent (kVA), active (kW) and reactive (kVar) power flow from node i to node j
s	Root node/source node
S_{branch}^{max}	Maximum rating or capacity of branch (kVA)
S_{branch}	Complex power flowing through the branch (kVA)
$SLoss_{e_{ij}}, PLoss_{e_{ij}}, QLoss_{e_{ij}}$	Apparent (kVA), active (kW) and reactive (kVar) power loss of the edge connecting the nodes i and j
$SLoss_{Total}$	Total power loss of the network (kVA)
S, P, Q	Apparent power (kVA), active power (kW) and reactive power (kVar)
SS_C	Installed capacity of a secondary substation (kVA)
$SS_{C,avg}$	Average installed capacity of a secondary substation in the network (kVA)
$Ssupplied_{Total}$	Total supplied power to the network (kVA)
std	Standard deviation
$Total\ area$	Total supply area of a network (km^2)
V	Line to line Voltage (V)
V_d	Voltage drop occurs in a distribution line segment (V)
V_{d_Total}	Total voltage drop along a distribution feeder (V)

V_i, V_j	Voltage at nodes i and j (V)
y	Random variable used in the kernel density estimation
$\overline{y}$	Mean/expected value of the random variable *y*
$z(i,j)$	Per km impedance of the distribution line connecting nodes i and j
Z, R, X	Impedance, resistance and reactance (Ohms)
z, r, x	Per km impedance, resistance and reactance of the conductors (Ohms/km)
φ	Angle between voltage and current (degrees)

Parameters and variables in Chapter 5

$f_i(k)$	PMF of the depth dependent degree distribution at level i
$f(l,\alpha)$	PDF of the depth dependent edge length distributions
$f(n)$	PMF of the distribution of nodes among the levels of a network
i	Level of the network
k_{max}	Maximum node degree in the network
k_s	degree of the source node *s*
LD	load density (MVA/km^2)
L_{max}	Maximum feeder length of the network (km)
$l_{\min}, l_{\max}$	Minimum and maximum edge lengths observed at each level of the network
M_i	Number of nodes in the network up to level i
N	Total number of nodes in a network
n_{leaf}	Number of leaf nodes in a network
p	Path from source node to a leaf node along the feeders of the network
pf	Power factor
p_{ik}	Probability that a degree of a node in the network level i is equal to k

$PLoss_{path\_p}$	Active power loss occurs in the path p (kW)
q_i	Probability that a given node in the network is in level i
r_{supply_max}	Maximum supply radius of a primary substation (km)
Std.Err.	Standard error of the exponential distribution compared to the mean of the actual dataset
t	Network type
V_{d_max}	Maximum allowable voltage drop of the network (V)
α	Mean of the exponential distribution
σ_{approx}	Approximate load density (MW/km^2)

Parameters and variables in Chapter 6

$\mathrm{C}_{\mathrm{loss}}^{\mathrm{ref}}$	Annual energy loss cost of the distribution network in the reference case, where no SOP is installed (£k)
$\mathrm{F}^{\mathrm{ref}}$	Annual cost of the distribution network in the reference case, where no SOP is installed (£k)
$\mathrm{F}^{\mathrm{SOP}}$	Annual cost of the distribution network with one SOP installed (£k)
P_{C1}, Q_{C1}	Active (kW) and reactive (kVar) power provided by VSC1
P_{C2}, Q_{C2}	Active (kW) and reactive (kVar) power provided by VSC2
C_{inv}	Investment cost of SOP (£k)
C_{loss}	Annual energy loss cost of a distribution network (£k)
C_{ope}	Operational cost of SOP (£k)
I_k	Current flow through branch k (A)
I_k^{max}	Maximum allowed current of branch k (A)
N_S	Set of all scenarios considered over a year
N_{branch}	Total number of branches
N_{bus}	Total number of buses in a network.
$P_{C1,sce}$, $P_{C2,sce}$	Active power exchanged through the converters 1 and 2 of an SOP at scenario sce (kW)

P_i, Q_i	Active (kW) and reactive (kVar) power flowing from bus i to bus i +1
$P_{load(i)}$, $Q_{load(i)}$	Active (kW) and reactive (kVar) power demand at bus i
$P_{loss(i,i+1)}$, $Q_{loss(i,i+1)}$	Active (kW) and reactive (kVar) power loss of the branch connecting buses i and $i+1$
$P_{loss,sce}$	Power loss of the network per unit time at scenario sce (kW)
S_{SOP}	Capacity of an SOP (MVA)
S_{module}	Minimum capacity of the basic power electronic module in an SOP (kVA)
V_i	Voltage at bus i (V)
V^{min}, V^{max}	Minimum and maximum bus voltage limits (V)
$b_{i,j}$	Binary variable that indicates if the branch between buses i and j is equipped with an SOP
c_{SOP}	Investment cost per unit capacity of an SOP (£/kVA)
c_{ele}	Electricity price per kWh (£/kWh)
p_{sce}	Probability corresponding to scenario sce
r_i, x_i	Resistance and reactance of the branch connecting buses i and $i+1$ (Ohms)
r_{inv}	Discount factor for the investment cost of SOP
r_{ope}	Discount factor for the operational cost of SOP
F	Objective function
$f(x)$	Continuous probability density function of variable x
m	Number of basic power electronic modules
sce	Scenario number
y	Device economical service time of SOP (years)
η	Loss coefficient

Chapter 1: Introduction

1.1 Background

The electrical power system is one of the most critical infrastructures of a country. Traditional power systems usually consist of large centralized power generation units connected to the High Voltage (HV) grid, which produce electric power mostly by burning fossil fuels. But, the modern power systems have become more decentralized due to the large penetration levels of renewable power generation (e.g. Solar photovoltaics (PV), wind, hydro and tidal power) and other distributed energy resources (DER) (e.g. Battery storage, Electric vehicles (EV)), which are mainly connected to the power grid at Medium Voltage (MV) and Low Voltage (LV) levels. These transformations are mainly driven by:

(i) policy goals to achieve energy and climate change objectives by reducing greenhouse gas emissions [1], [2], and

(ii) the challenges imposed by depletion of fossil fuel reserves.

Technologies such as renewable power generation and DERs that help to reduce greenhouse gas emissions belong to Low Carbon Technologies (LCTs). With the increasing integration levels of LCTs at the MV and LV levels, electricity distribution networks are of increasing importance in modern power systems.

1.2 Analysing uncertainties in electricity distribution networks

1.2.1 Uncertainties in electricity distribution networks

The effective integration of the LCTs into the existing distribution networks requires significant changes of the technical, regulatory and commercial arrangements in the current systems. These changes have led to many techno-socio-economic uncertainties in the electricity distribution systems. In general, the various sources of uncertainties in electrical power systems can be summarised as shown in the Figure 1.1.

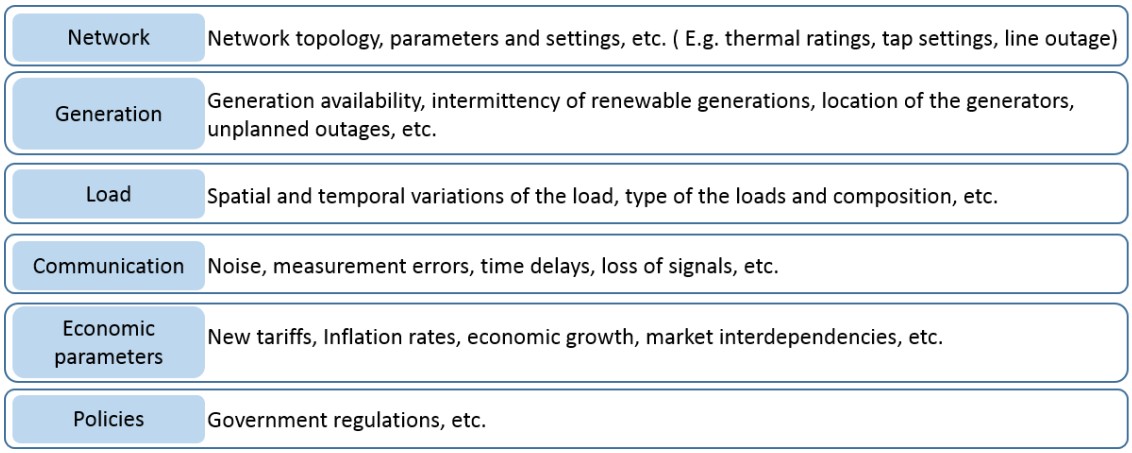


Figure 1.1: Sources of uncertainties in electrical power networks.

Due to the integration of new LCTs, significant levels of technical uncertainties are introduced to the electricity distribution networks by the changes in network configurations, variations in electricity generation and load demands. In addition, socio-economic uncertainties such as, uncertainties in policy commitment and support for decarbonisation in energy sectors, uncertainties in economic and financial viability, future costs, market conditions and public acceptance of the new LCTs are also associated with the integration of LCTs in the distribution networks [3].

Therefore, it is important to identify effective tools and methodologies to understand (i.e. analyse and quantify), manage and mitigate these uncertainties associated with LCTs in the distribution networks for operational/strategic decision-making and policy support.

1.2.2 Methods of uncertainty assessment and modelling in electricity distribution networks

Much research has been carried out worldwide to model, analyse and quantify the uncertainties associated with LCTs on electricity distribution networks. The methods used in previous research can be categorised as shown in Figure 1.2.

Many previous studies on the impact assessment and modelling of LCTs have been deterministic in approach, capturing the uncertainties using scenario or sensitivity-based analysis [4], [5], [6]. These deterministic methods do not consider the probabilistic nature of the system behaviour or parameters and provide deterministic solutions in the presence of uncertainties. The worst-case scenario analysis is a commonly used deterministic method for making decisions in the face of uncertainties.

In addition to the deterministic studies, a number of previous research on distribution networks have considered the probabilistic behaviour of system parameters, e.g. variation of demand and generation, probability of component failures, etc. [7], [8], [9]. The probabilities of the occurrence of an event in the distribution networks are recognised using probabilistic studies in order to support decision making activities in the distribution systems.

On the other hand, data driven modelling techniques are gaining popularity to study uncertainties in many scientific disciplines in the recent years. Through collecting, summarising, analysing and interpreting a large amount of numerical data, rules can be derived to predict the behaviour of real systems in the face of uncertainties.

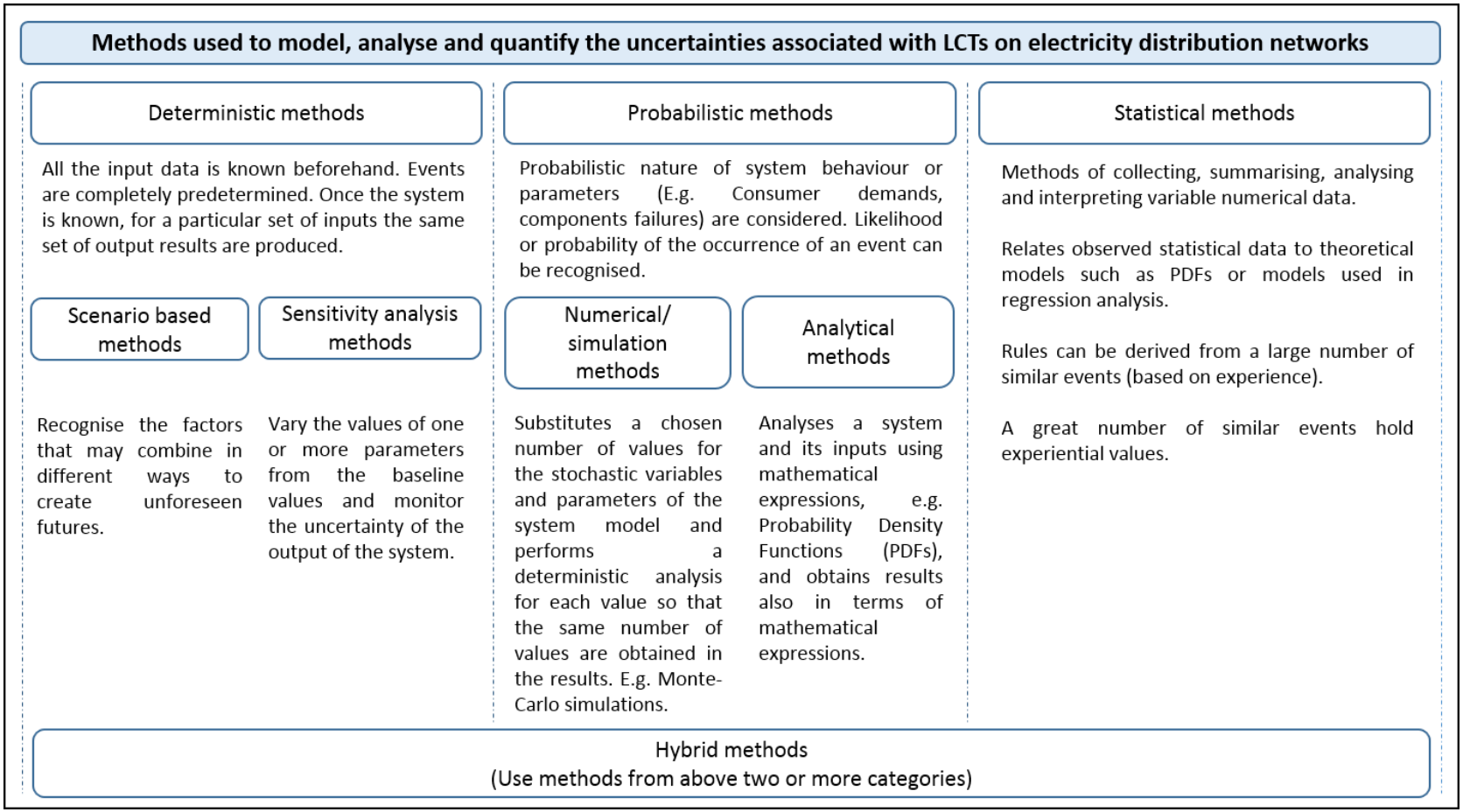


Figure 1.2: Methods used to model, analyse and quantify the uncertainties associated with LCTs on electricity distribution networks.

So far, these kinds of data driven models with relevant to the impact assessment studies of LCTs are largely used in the applications such as, predicting intermittent renewable generation, recognising EV charging patterns and demand forecasting.

The hybrid methods, which use methods from above two or more categories are usually more comprehensive in approach and better decision support is provided than the individual methods.

1.2.3 Challenges associated with existing uncertainty assessment and modelling methods and research gaps

A major challenge for the government and industry is to assess the impact of new LCTs on electrical power systems and make robust decision in the face of significant uncertainties. For example, it is important for policy makers to be able to characterise and quantify how differently the urban and rural networks perform with different integration levels of a new LCT.

However, most of the above deterministic and probabilistic studies are conducted on real network samples [4], [6], standard synthetic networks such as the IEEE test cases [7], [10] or other representative test networks [11]. As a result, most reported analysis in the literature is only useful for evaluating a specific test case and conclusions made from such studies have limited applicability to other networks.

A few researchers have identified the need of data driven/statistical studies to support the decision making activities based on impact assessment studies on a large number of realistic electricity networks [12], [13], [14], [15]. This kind of decision making approach on uncertainty assessment studies of power networks, is more robust compared to the decision-making approaches based on the case study networks.

For example, a large-scale network model based on fractal generation is used to generate LV and MV networks in [14], [15]. This fractal networks generator has been developed using a bottom-up approach where, the network generation has done by first, generating the realistic consumer distributions and then, by connecting the consumer locations to obtain the network topologies. The networks generated by the fractal generating tool do not correspond to any real networks but are capable to realistically mimic typical network characteristics of distribution networks of Great Britain (GB), thus allowing more general and strategic conclusions to be drawn with respect to case studies on specific networks. In [12] and [13], the impact of PV generation and EV integration is assessed on a complete distribution system of a single utility in New Zealand. Nevertheless, each distribution network is different, this kind of studies help to identify the common behaviours of different types (urban, sub-urban and rural) of distribution networks.

However, it is challenging to obtain detailed data from a large number of real networks, to carry out these simulation studies. In most cases access to the real-world network data is very limited. Many researchers spend a large amount of time searching for real network data and cleaning the data. On the other hand, the number of available representative test networks such as the IEEE test cases and the test cases from UK generic distribution system (UKGDS), are also limited to carry out statistical studies on different types of distribution networks.

Other than the aforementioned fractal networks generator for GB electricity distribution networks, comprehensive statistical network generating tools for distribution networks were not found in the open literature. The characteristics of distribution networks are different from country to country and region to region. Hence more statistical network generation tools supported by large amount of real world network data are required, to support strategic and robust decision making on the distribution networks studies.

1.3 Objectives, scope and the idea of the research

The aim of this thesis is to present the rationale and development of a network modelling tool with the ability of generating a large number of random, realistic models of electricity distribution networks and demonstrate how such a tool can be used to analyse the risks and benefits associated with LCTs in order to support decision making. The overall concept of the tool is illustrated in Figure 1.3.

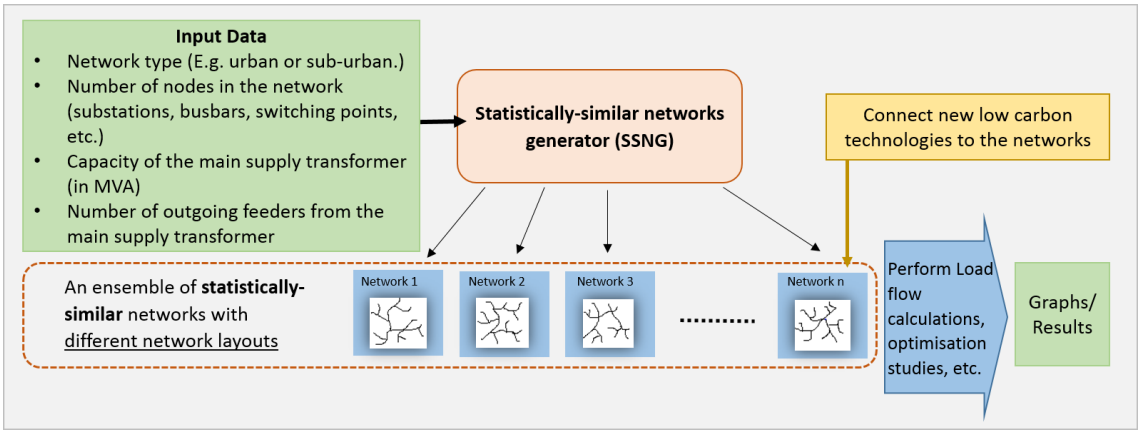


Figure 1.3: The concept of the network assessment tool.

Statistically-Similar Networks Generator (SSNG) is the main element of the proposed tool. With a give set of input data, the SSNG is capable of generating a user defined number of 'statistically-similar' networks. Statistically-similar distribution networks are a set networks with a similar set of topological and electrical properties as defined by the user with some given values or ranges of values.

The generation of multiple representations for the same real-world network (i.e. in this case, electrical power network) has always been problematic in network modelling. A single representation might not capture all the diverse characteristics of a real network. Therefore, the ability to generate statistically-similar many networks to represent an electricity distribution network is one of the key features of the proposed tool.

The network modelling and simulation tool proposed in this thesis is named 'A statistical assessment tool for distribution networks' for the reason that, the conclusions about the impacts of LCTs on distribution networks are now derived from the statistical studies conducted on a large number of statistically-similar networks.

The main difference between the fractal networks generator and the networks generator presented in this thesis is that, the later uses a top-down approach for the network generation. In the top down approach, the networks generation is based on the graph related topological properties and electrical properties of the real-world networks without going into the details of the actual/realistic geographical location/distribution of the individual consumers.

In order to develop the aforementioned network assessment tool, this thesis set out to achieve the following objectives.

(i) To review the statistical studies on power networks and to identify the available statistical network modelling tools, compare them and identify their limitations.
(ii) To investigate the topological and electrical properties of real world electricity distribution networks and identify and quantify the key features that characterise different types of distribution networks.
(iii) To develop a statistically-similar networks generator for electricity distribution networks.
(iv) To demonstrate the application of the SSNG tool - Analyse the impacts of Soft Open Points (SOPs) on electricity distribution networks with variable levels of DG penetration.

1.4 Thesis outline

The structure of the thesis follows the research objectives and has seven main chapters. The structure of the thesis is shown in Figure 1.4.

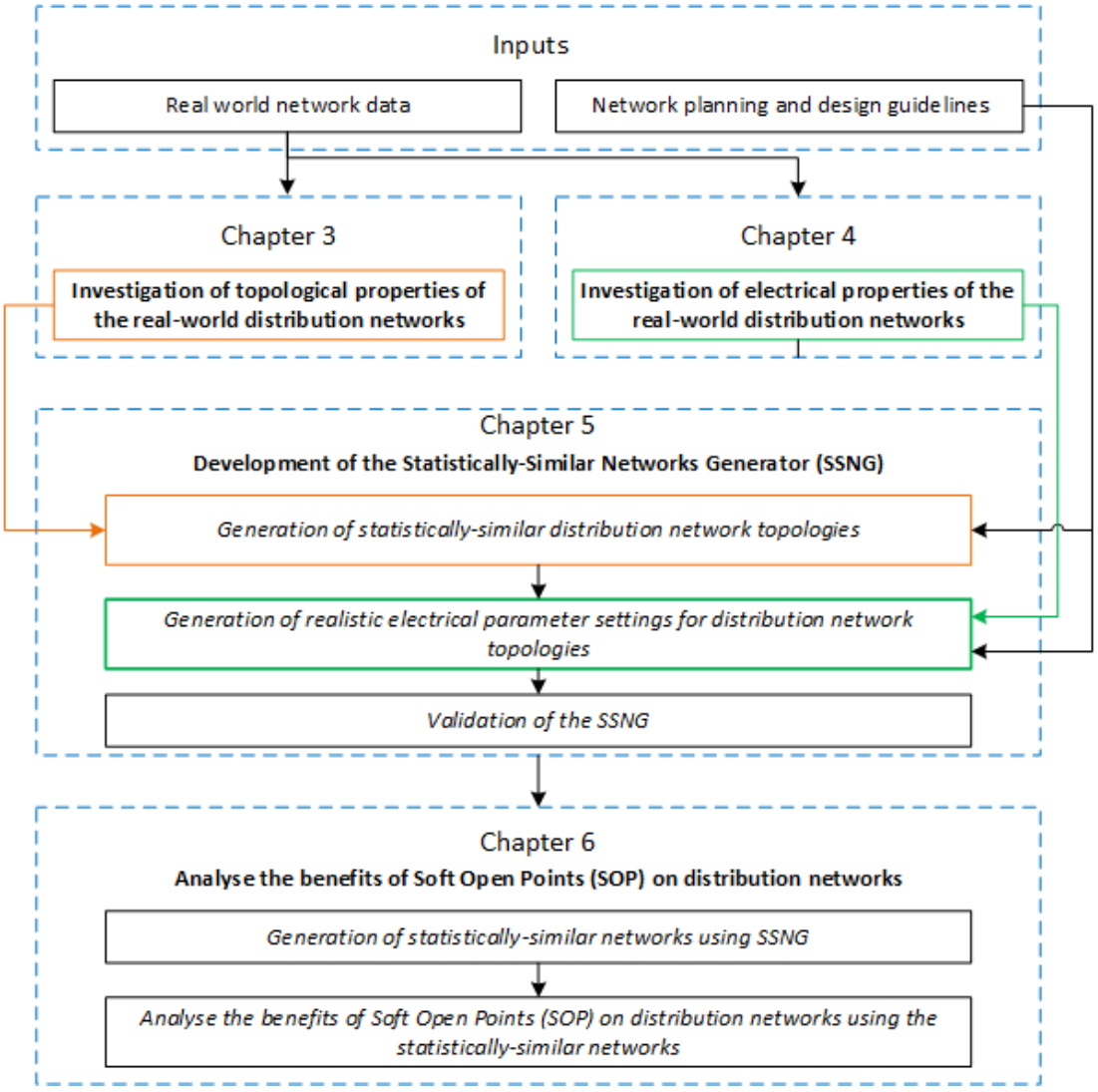


Figure 1.4: The thesis structure.

Chapter 2 first provides a background on the different type of decisions to be made in power systems planning, design and operation relative to the level of decision stakes and the level of uncertainties. A review of the literature on impact assessment studies of LCTs on distribution networks including the different methods of analysing uncertainties in the electrical power

networks is presented. The statistical studies conducted on electrical power networks and various network models and network generating tools available for modelling the electrical power networks are also reviewed.

In **Chapter 3,** an investigation of the topological properties of the real-world networks is conducted using the network data which was collected from China, covering urban and sub-urban areas. The key topological properties that characterize different types (urban and sub-urban) of distribution networks are identified and quantified.

In **Chapter 4,** electrical properties of the real-world networks are studied using the same network data which was collected from China, covering urban and sub-urban areas. The key electrical properties that characterize different types (urban and sub-urban) of distribution networks were identified and quantified.

In **Chapter 5,** the development process of a Statically-Similar Networks Generator is presented. The input parameters required for the generation of realistic distribution networks are taken from the investigations of topological and electrical properties which were conducted in the Chapters 3 and 4 respectively. The guidelines for planning and design of the real-world distribution networks are also employed in the process of SSNG development. First, the methodology to generate statistically similar network topologies is presented. Then the procedure to generate realistic electrical parameter settings for the distribution network topologies is presented. Validation of the SSNG model is done by comparing the performance and properties of the generated statistically-similar networks with a real-world network sample.

Finally, in **Chapter 6**, the impacts of Soft Open Points (SOPs) on the distribution networks are analysed and quantified using a set of statistically-similar distribution networks that are

generated by the SSNG. General conclusions are derived on the impact of SOPs on the selected type of distribution network through a statistical study.

The conclusions and possible future work are summarised in **Chapter 7**. The limitations of the present work are also discussed.

1.5 Contributions of the research and publications

1.5.1 Contributions of the research

The contributions of this thesis are as follows.

- An investigation of the topological properties of real-world electricity distribution networks using the techniques in complex networks theory and graph theory is presented. The key topological properties that characterise different types of electricity distribution networks were identified and quantified. A novel depth dependent approach was developed to investigate the network topologies.

- An investigation of electrical properties of real-world electricity distribution networks is presented and the key electrical properties that characterise different types of electricity distribution networks were identified and quantified. A novel depth dependent approach was developed to investigate the electrical properties of the distribution networks.

- A data driven model based statistically-similar networks generator (SSNG) for electricity distribution networks was developed. Real-world network properties as characterised by the above topological and electrical investigations and the corresponding network planning and design guidelines were used in the development of the SSNG.

- A statistical analysis of the impact of Soft Open Points (SOPs) on electricity distribution networks with variable distributed generation penetration was presented on a set of statistical-similar distribution network models. General conclusions were derived about the impact of the SOPs on the type of electricity distribution networks under study.

1.5.2 Publications

The following papers are published with relevant to the work described in this thesis.

- Sathsara Abeysinghe, Jianzhong Wu, Mahesh Sooriyabandara, Muditha Abeysekera, Tao Xu, Chengshan Wang "Topological properties of electricity distribution networks", In Applied Energy, 2017, ISSN 0306-2619.
- Sathsara Abeysinghe, Jianzhong Wu, and Mahesh Sooriyabandara, "Statistical Assessment Tool for Electricity Distribution Networks," Energy Procedia, Volume 105, May 2017, Pages 2595 – 2600.
- Sathsara Abeysinghe, Silviu Nistor, Jianzhong Wu, Mahesh Sooriyabandara, "*Impact of Electrolysis on the Connection of Distributed Generation*", Energy Procedia, Volume 75, August 2015, Pages 1159-1164.

Chapter 2: Literature review

2.1 Introduction

Dealing with uncertainties in the electrical power networks has always been one of the main concerns of decision makers including policy makers, system operators and industry stakeholders. This chapter first provides a background on the different type of decisions to be made in power systems planning, design and operation relative to the level of decision stakes and the level of uncertainties.

With a large number of LCTs connecting at the MV and LV levels, making robust decisions under uncertainties at the distribution levels of the power networks has become more challenging. Thus, the present study has focused on investigating the methods and tools available to support robust decision making on the impacts of LCTs on distribution networks. A review of the literature on impact assessment studies of LCTs on distribution networks is presented.

The importance of statistical studies on power networks and statistical modelling of power networks to support robust decision making have been identified. The existing literature on analyzing statistical properties of power networks and various methods and tools available for modelling the electrical power networks are also reviewed.

2.2 Decision making in electrical power systems

In the context of power system planning, design and operation, different types of decisions are involved. Figure 2.1 illustrates the use of Funtowicz and Ravetz (1990) model adapted to understand uncertainties in the energy systems [16]. The model classifies type of decisions to be made in problem solving relative to the level of decision stakes and the level of uncertainty.

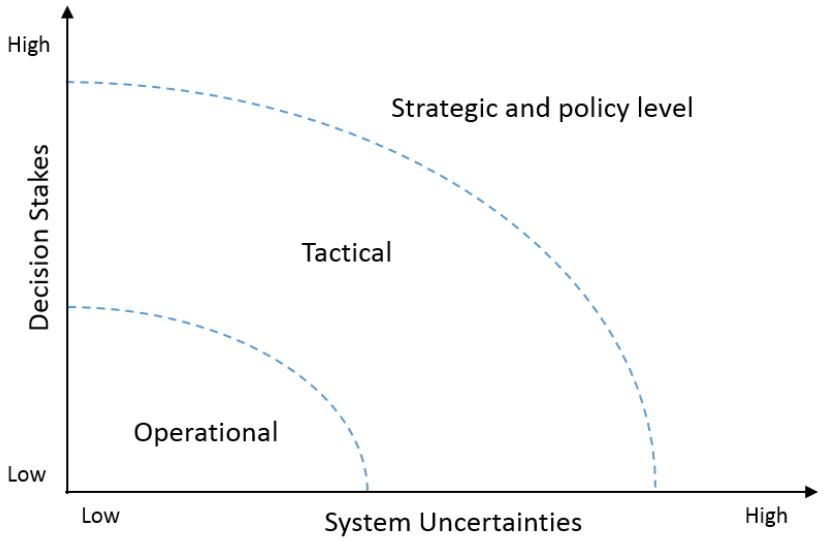


Figure 2.1: Types of decisions to be made in power system studies relative to the level of decision stakes and the level of uncertainty [16].

Different types of decisions involved in power systems studies are explained below.

- **Operational decisions**

Low system uncertainties, describe a situation where uncertainties are at the plant/network level and low decision stakes (operational decisions) describe decisions that are usually applicable to a single operator [16]. In operational level decisions there is little concern about the long-term impact and the concern is mostly on the short-term operation of a given system. For instance, decisions about how much energy to be produced by the system to meet the end user demand is an operational decision. Such operational level decisions are supported by optimization and simulation models where the given system is analyzed/optimized in various possible configurations and under different scenarios [17].

- **Tactical decisions**

Tactical decisions are mostly medium-term decisions that can prepare the system for strategic changes. For example, if the energy demand of a region is expected to grow significant investments on the power supply system will be needed. Integration of new renewable

generators can be a potential solution. Such tactical decisions require an analysis of what implications each relevant alternative may have on the local power system and on the local community [17]. These decisions exist at the methodological level and often require more than one model to assess the uncertainties and several decision makers with multiple, conflicting criteria are involved. As a result, tactical decisions can be informed, but cannot be solved, by operational level techniques [16].

- **Strategic and policy level decisions**

Strategic and policy level decision making is a complex process and encompasses high levels of uncertainties over medium and long term, inputs from various stakeholders and massive investments. In general, it is difficult to structure and model the all these factors that may involve at this level of the decision making. However, all possible implications that may affect the future of society and the environment must be considered in the process. Tactical and operational level analysis can be used to inform strategic and policy level decisions up to some extent [16], [17].

With high penetration levels of LCTs at the distribution levels of the power networks, it is important to identify the methods and tools that can support robust decisions making in the power system studies at operational, tactical, strategic and policy levels.

2.3 Impact assessment studies of LCTs on distribution networks - A review

Table 2.1 provides a literature review of the uncertainty assessment and modelling studies of LCTs that have been conducted on electricity distribution networks. The details of the uncertainties analyzed/optimized, the number of networks and the type of the networks used in the study, the methods used for the uncertainty assessment and modelling are compared.

In general, the focus of these studies has been on the analysis of technical and economic challenges of the integration of new LCTs, evaluation of hosting capacity of DG, investigation of negative impact mitigation options of LCTs and sizing of the devices that will be used for negative impact mitigation in the distribution networks.

According to the review in Table 2.1, most of the studies in literature on the impact assessment of LCTs have focused on one or few real or synthetic network samples. A few studies which have been conducted on a very large amount of realistic/real-world network samples are identified [12], [13].

Table 2.1: A literature review of the uncertainty assessment and modelling studies of LCTs on distribution networks

Ref.	LCT	Uncertainties	Case study networks used	Methods used for uncertainty assessment and modelling
[4]	Solar PV	Voltage magnitude, unbalance	A residential LV network in Malaysia	A deterministic study. Sensitivity analysis with varying PV penetration levels is used.
[7]	Solar PV	Voltage magnitude	IEEE 33-node radial system	A probabilistic study. Both analytical techniques and numerical methods are used to solve probabilistic load flow.
[11]	Solar PV	Power Quality, under-voltage, over-voltage, Unbalance	Four representative (LV rural, semi-rural, urban and city) feeder types from Flanders, Belgium	A deterministic study. Scenarios have been designed considering feeder topologies, load and generation profiles, PV penetration level, PV-size statistics and the PV location.
[12]	Solar PV	Over-voltage and under-voltage problems	Entire LV network data from a single utility	A data driven/statistical study. Monte Carlo type simulations are performed on a large number of realistic LV distribution networks of different types (rural, urban)
[8]	Solar PV and PHEV charging stations	Energy losses, reactive power support to the grid, voltage magnitude, peak load reduction	An 18-bus test feeder	A probabilistic study. PV output is generated from a probabilistic solar irradiance model.
[10]	PHEV	Power losses, main grid load factor and voltage magnitude	IEEE 34-node test feeder	Both deterministic and probabilistic methods are used.
[5]	EV	Distribution network investment, energy losses	Two large-scale distribution networks with several voltage levels	A deterministic study. Three scenarios with 35%, 51%, and 62% of EV penetration are used.
[13]	EV	Over-voltage and under-voltage problems, overloading of distribution lines	Entire LV network from a single utility	A data driven/statistical study. Monte Carlo type simulations are performed on a large number of realistic LV distribution networks of different types (rural, urban)

[9]	Wind	Limits on the distribution feeders' voltages and currents and costs of energy	A 69-bus radial distribution system and a 152-bus radial distribution system	A probabilistic study. Probabilistic behaviour of wind generation and consumer load is considered.
[18]	Energy Storage Systems (ESS), wind power	Levelised production cost of energy, wind energy curtailment	A section of the South Wales distribution network in UK and one UK Generic Distribution System (UKGDS)	A deterministic study. Scenarios are chosen depending on windfarm capacity and ESS capacity allocation.
[6]	DG and electrolytic hydrogen production	Voltage magnitude, line loading, wind energy curtailment	An actual distribution network sample from the UK	A deterministic study. Five scenarios were considered depending on different levels of DG connection and control strategies.
[14]	CHP and electric HP	Change in peak demand, thermal and voltage limit violations	Two UK representative urban and rural LV networks generated from a fractal network generator	A deterministic study. Scenario studies have been carried out by modelling different penetration levels of consumer points with DCHPs or electrical HPs.
[19]	Microgrids with renewable generators	Voltage magnitude, line losses	Sample LV distribution network with two feeders and 10 consumer nodes	A deterministic study. Scenario studies have been carried out with different demand profiles and solar irradiance profiles.

Depending on the methods used for uncertainty assessment and modelling and depending on whether the study is carried on a large number of real/realistic networks, the placement of all above previous studies on the Funtowicz and Ravetz (1990) model adapted to understand uncertainties in the energy systems is shown in Figure 2.2.

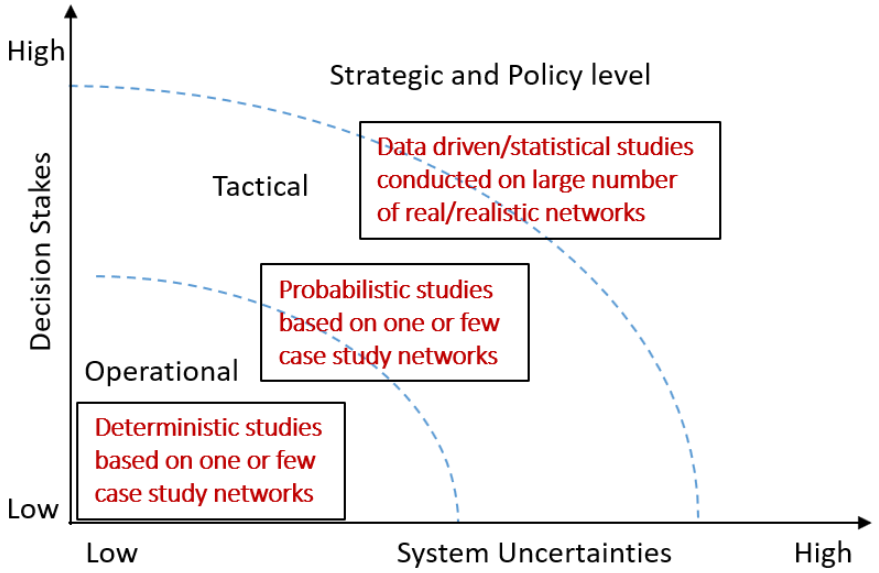


Figure 2.2: Placement of the uncertainty assessment and modelling methods of LCTs on distribution networks according to the level of decision support provided by the study.

- **Deterministic studies**

According to the literature review on impact assessment studies of LCTs, in the deterministic type studies which are based on one or few case study networks, all the input data for the analysis is known beforehand and the events are completely predetermined. Therefore, for a particular set of inputs the same set of output results are produced. For example, some researchers have performed time-series simulations using historical load and generation profiles to determine what is running in each time period. This type of studies provides information on the variations of voltage, current and power levels throughout simulation period in a specific distribution system and the conclusions derived from these studies can be used to support operational level decisions in that specific electrical distribution system. For instance, results

from the deterministic study in reference [4] indicated that the LV network used in the study can accommodate high level PV (>100%) penetration without causing any problem. However, it is important to mention that the impact of PV is very location and network dependent.

Since, the deterministic studies do not take into account the probabilistic behaviour of the load, generation and other system parameters robust decisions in the tactical, strategic and policy levels cannot be derived from such studies. However, the deterministic studies based on scenarios and sensitivity analysis studies can inform the tactical and strategic level decisions on the factors that may combine in different ways to create unforeseen impacts on the distribution networks.

- **Probabilistic studies**

Although, the probabilistic studies are able to deal with the uncertain parameters of the distribution systems, most of such studies in the literature are based on the case study networks. Hence, the conclusions derived from those studies have limited applicability to the other networks. Therefore, robust and generalized conclusions required for strategic and policy level decision making cannot be derived by solely depending on the results of these case study networks-based probabilistic studies. However, the conclusions made from such probabilistic studies can support operational levels decision making on the specific distribution network used in the study. For instance, in the probabilistic load flow study in reference [7] it has been recognized that by connecting PV generators to certain nodes the voltage profile in the given distribution network can be improved.

- **Data driven/statistical studies**

In general, data driven/statistical studies relate the observed statistical data to theoretical models such as PDFs or models used in regression analysis. Rules are derived from a large

number of similar events or based on the experience obtained by conducting data driven studies.

In this thesis the impact assessment studies of LCTs conducted on a large number of distribution networks are given much attention in the category of data driven or statistically driven studies. The abovementioned deterministic and probabilistic type analysis are often conducted on a large number of networks and the statistics of the results of many similar networks are used derive the conclusions of such data driven studies. Therefore, it was identified that the conclusions made from these studies are more robust and generalized than the aforementioned one or few case study networks-based studies. For example, studies [12] and [13] claim to provide information on the proportion of the distribution network that may experience problems by integrating LCTs and highlights the importance of undertaking LCT impact studies on a large number of networks to come up with robust and generalized conclusions to support strategic and policy level decision making.

The challenges and limitations to conduct such statistical studies have been identified [20]. They are,

(i) Difficulty to obtain realistic network data from utilities and the amount of time and efforts need to clean the data.

(ii) Limited availability of reference test networks.

(iii) Limited availability of open source tools for the generation of representative test cases of the real networks.

2.4 Analyzing statistical properties of electrical power networks - A review

Followed by the review of impact assessment studies of LCTs, the need for generating random-realistic models that can effectively capture key topological and electrical characteristics of real-world networks in aiding statistical analysis of the power networks is identified [12], [13], [14].

However, it is often difficult to produce a large number of random-realistic models of the real-world networks including the electric power networks. Identifying and quantifying the key statistical properties that characterize different types of networks is a fundamental requirement when developing such random-realistic models of real networks.

2.4.1 A summary of the related previous research

The statistical properties of the power networks have been studied by many researchers in the past. The existing literature on analyzing statistical properties of power networks have been reviewed in this section. Statistical properties of the power networks are usually studied under two categories; topological and electrical properties. Table 2.2 summarizes some of the related studies. The type of the networks, voltage levels, number of networks used and the topological and electrical properties that are investigated in each study are compared.

Table 2.2: A literature review on analysing the statistical properties of power networks [21].

Ref.	Network type	Voltage level	Number of networks analyzed	Node degree related statistics	Average shortest-path length	Clustering coefficient	Betweenness	Fractal properties	Electrical distance/ impedance distribution
[22]	Real	HV	1 real grid (Italian grid)	✓					
[23]	Real and synthetic	HV	1 real, 1 random grid & 1 IEEE test network	✓					
[24]	Real	HV	1 real grid (North American)	✓					
[20]	Real and synthetic	HV	4 IEEE test networks and 2 real networks	✓	✓				✓
[25]	Real and synthetic	HV	1 real and 1 synthetic	✓	✓	✓			✓
[26]	Real and synthetic	HV	2 IEEE test cases and 1 real grid				✓		✓
[27]	Real	HV	1 real grid (European power grid)	✓	✓	✓			
[28]	Real	HV	1 real grid (North American)	✓					
[29]	Real	HV	3 real grids (Italian, French and Spanish grids)						
[30]	Synthetic	HV	1 IEEE test case		✓				✓
[21]	Real	MV/LV	Northern Ireland network data	✓	✓	✓	✓		✓
[31]	Real	MV/LV	2 LV networks and 2 MV feeders					✓	

In the past, the interest of studying the statistical properties, primarily the topological properties of electrical power grid was mainly led by the major blackouts happened in North America [28], [24] , Italy [22], Europe [27] and few other power grids [32]. After these large-scale blackouts happened worldwide, researchers were seeking solutions for improving security and reliability of the power grid from different perspectives. Advances in statistical physics and complex network theory together with graph theory applications have developed new areas of interest in vulnerability assessment in power systems [33]. The connection between the structure of the HV power grid and the risk probabilities are often analyzed.

As seen from the review in Table 2.2, majority of the statistical studies on the power networks are carried out in the HV networks. The number of statistical studies on the MV and LV networks are very limited. To the best of the author's knowledge only one comprehensive topological investigation of the MV and LV networks was found in the literature which has used data from a large number of real networks. The study has focused on a methodology of integrating topological metrics with economic factors and the topological characteristics of MV and LV networks are compared [21]. Other than the aforementioned MV and LV network study, the statistical investigations of power networks are conducted on a limited number of networks in most of the above previous research.

Complex networks analysis techniques together with graph theory applications have been used in majority of the studies to analyze the statistical (both topological and electrical) properties of the power networks.

In addition, fractal properties of the power networks are also investigated under the topological properties [31].

2.4.2 Frequently analyzed statistical properties of power networks

Most commonly investigated statistical properties of the power networks are summarized in this section. In order to extract the graph related network properties (i.e. node degree, average shortest-path length, clustering coefficient, betweeness and distribution of impedances related statistical measures), the electrical power networks have been modelled as graphs with electrical substations and consumer load points representing the nodes of the graphs and transmission and distribution lines representing the edges of the graphs.

- **Node degree related statistical measures**

The number of edges incident with a node in the network is called the degree of that node. Different nodes in the network can have different node degrees. Degree distribution is the probability distribution of the degrees of nodes over the whole network [33]. Node degree related statistical measures such as the degree distribution, average degree of the network and maximum node degree of the networks are widely studied in the previous research in Table 2.2.

Researches have observed that the HV electrical transmission grids have heavy-tailed nodes degree distributions. A network is said to be 'scale-free' when the nodes degree distribution of the network follows a power law, resulting in few nodes having many edges and many nodes having few edges. In vulnerability assessment studies of HV networks, it has been identified that networks with scale free property are highly resistant to accidental failures but are highly vulnerable to deliberate attacks [34].

- **Average shortest-path length**

The average path length is defined as the average length along the shortest paths for all possible pairs of network nodes [33]. It has been widely used as a measure of the efficiency of

information or mass transport on a network. Higher average shortest path length implies network is almost in liner chain and lower characteristic path length shows the network is in a compact highly meshed form. Most of the previous studies under review have studied the average shortest-path length property of the networks.

- **Clustering coefficient**

In graph theory, a clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. Evidence suggests that in most real-world networks, and in particular social networks, nodes tend to create tightly knit groups [33]. Therefore, clustering coefficient of the power networks have been studied in some of the above previous research to capture the local cohesiveness of the networks.

- **Betweenness**

Betweenness describes the importance of a node with respect to shortest paths in the graph. This is very important to identify critical components of the power grid [21], [22], [35]. For a given node, betweenness, is defined as the number of shortest paths that traverse that node. Considering the betweenness of a node as a random variable it is possible to obtain the corresponding probability distribution. Betweenness properties have mostly been analysed in the vulnerability assessment studies of the power networks.

- **Fractal properties**

Historically, Benoit Mandelbrot first introduced the concept of fractals. In general, a fractal can be described as a rough or fragmented geometric shape that can be subdivided in parts, each of which is (at least approximately) a reduced/size copy of the whole [36]. According to this definition, a same repeating pattern can be seen at every scale of a fractal. Fern leaf is a simple

example for a fractal. In the fern leaf shown in the Figure 2.3 it can be noticed that every little leaf is part of the bigger one and has the same shape as the whole fern leaf. This property of the fractals is called the self-similarity. A fractal dimension is a statistical index for characterizing fractal patterns by quantifying their complexity as a ratio of the change in detail to the change in scale [36].

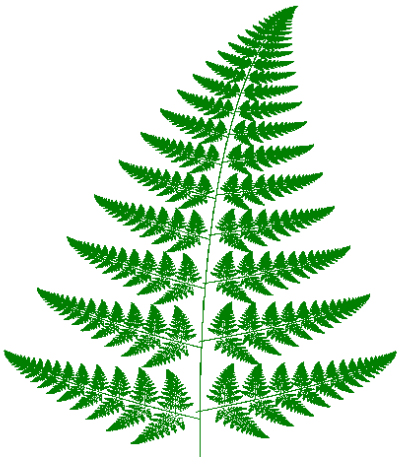


Figure 2.3: Fern leaf-an example for fractals.

Some researchers have observed that power networks have fractal like structures with self-repeating patterns at all length scales [15], [31]. In reference [31], an analysis of the fractal properties of two LV networks and two MV feeders is presented. Through this study it was identified that the four networks under the study have fractal characteristics but four different fractal dimensions.

- **Electrical distance/impedance distribution**

Investigations of electrical properties of the power networks in the literature mainly refer to the investigations related to the distribution of transmission and distribution line impedances of the power networks. Weighted graphs of the power networks with weights of the edges representing the impedance of the transmission or distribution lines are used to derive the

electrical properties of the power networks in such studies. The concept of electrical distance has also been used when studying the electrical structures of the power networks [20], [21], [25]. The most common measure of electrical distance has been the absolute value of the inverse of the system admittance matrix (i.e. this is the same admittance matrix that is used in power flow analysis in electrical power system studies) [37].

2.5 Methods and tools for the modelling of electrical power networks - A review

Mathematical modelling of a real-world network means producing an artificial object that has similar characteristics to the real one [38]. Followed by the investigations of statistical properties of the power networks, various methods and tools that has been developed in the literature to model the electrical power networks are reviewed in this section.

It has been identified that real-world networks share some important properties, irrespective of their functionalities. A number of complex networks models have been developed in the past to model many different types of real-world networks (e.g. internet, telecommunication networks and social networks). Therefore, some of the researchers have investigated the suitability of the well-known complex networks models to describe the characteristics of the power networks. The relevant literature is summarized under the section of 'power grids as complex networks' [39].

As researchers started to pay attention on the fractal behaviour of real-world networks, various fractal models for real-world networks have been developed. Therefore, the fractal-based network models developed for the power networks are reviewed in this section.

One of the main goals of this thesis is to develop a statistical-similar networks generator for electricity distribution networks. The idea is to generate large numbers of random-realistic test

cases for various power network studies. Therefore, a review of the literature with a similar interest on statistical models/tools for the generation of random-realistic power networks is also presented.

2.5.1 Power grid as a complex network model

Table 2.3 provides a review of the literature on power grid as complex networks. The voltage level, type (i.e. real/synthetic) and the number of the networks under the study in each previous research are compared. The type of the complex networks models tested/used against the selected real-world or synthetic power networks in the previous studies are also summarized. Most of the related research was found in the recent survey conducted by Giuliano Andrea Pagani and Marco Aiello (2013) in the reference [39].

According to the review in Table 2.3, a majority of the previous studies on power grids as complex networks has been carried out on the HV networks. Vulnerability assessment of the power grids has been the primary objective of most of those HV level network studies. Only one relevant study that has been conducted on MV and LV networks was identified [40].

According to the literature, well-known and much studied complex networks models for testing the power networks are the random graph models, small-world networks models and scale-free network models. There are few other complex networks models which have also been tested for modelling the power networks (E.g. Forest fire graph model [41], Kronecker graph model [42], R-MAT network model [43]). However, these models are not discussed in detail for this thesis, as they were not found to be suitable so far for modelling the power networks in the existing literature. Also very limited literature exists on their suitability for modelling power networks [40].

Table 2.3: A literature review on power grids as complex networks.

Ref.	Voltage level	Real-world/synthetic networks used for the study	Complex networks model			
			Random graph models	Small-world network models	Scale-free network models	Other complex networks models
[25], [44]	HV	IEEE 300 bus network and the Eastern United States power grid	✓	✓	✓	Minimum distance random graphs
[28]	HV	North American eastern and western electric grids			✓	
[45]	HV	Western U.S. power grid and Nordic power grid	✓		✓	
[46]	HV	Shanghai power grid		✓		
[47]	HV	Power grids of China and America		✓		
[48]	HV	Sichuan Chongqing grid and Guangdong grid in China		✓		
[49]	HV	Power grids of the Nordic countries and western USA	✓		✓	
[50]	HV	IEEE test networks (14, 24, 30, 57, 118 and 300 buses)			✓	
[40]	MV and LV	Medium and low voltage power grids of the North Netherlands	✓	✓	✓	Forest fire graph, Kronecker graph, R-MAT graph

In reference [25], it has been identified that the minimum distance random graphs are suitable to model HV power networks. Since, the interest of this thesis is on methods of modeling MV and LV networks such models are not discussed here in detail.

All the different types of complex networks models exist in the literature are characterized by specific statistical properties. The statistical characteristics of the well-known complex networks models and their suitability to model the power networks as found in the literature are summarised below.

- **Random graph models**

In the late 1950s, Paul Erdὸs and Alfred Rényi introduced a random graph model and later it helped scientists to discover that most natural phenomena can be described in terms of random graphs. Since 1950s random networks have been studied for several decades. The network model that Erdὸs and Rényi introduced is now known as 'classical random network model' or 'Erdὸs-Rényi (ER) network model' where the random graph is built by picking each possible pair of nodes and connecting them with an edge with probability *p* [51].

In the references [25] and [44], the topological characteristics of real power grids and the corresponding random graphs models are compared. These studies concluded that the random graph models do not provide substantial utility for modeling HV power grids. In [45] and [49] structural vulnerability of real power grids are compared with the corresponding random graph models. It was identified that the power grids are more sensitive to attacks than the random graph models of those power grids. The study in [40], investigated the topological metrics that the MV and LV networks need to satisfy for efficiency, resilience and robustness with the help of various complex networks models including the random graph models.

- **Small-world network models**

The Small-world networks are characterized by high clustering and short average path lengths. This model was first introduced by Duncan J. Watts and Steven Strogatz in 1998. In the small-world networks there is the ability to reach any given point within the network in a fairly small number of steps relative to the network size. This small-world phenomenon is popularly known as six degrees of separation [52].

From the studies [25] and [49], it was identified that small-world networks are not suitable for modelling the Eastern United States power grid and the IEEE 300 bus network. However, small-world properties are observed in several Chinese power grids [46], [47], [48]. Performances of these Chinese power grids have been evaluated using the corresponding small-world network models.

- **Scale-free network models**

Scale free networks are characterized by power-law degree distributions [34]. Out of many different ways of developing networks with power-law degree distribution, the preferential attachment network model by Barabási and Albert has widely been used in the literature for real-world networks modelling. Preferential attachment means that the more connected a node is, the more likely it is to receive new edges. In other words, nodes with higher degree have stronger ability to grab new edges added to the network [53].

The scale-free nature of HV level power networks has been observed in a number of previous research. Therefore, in many vulnerability assessment studies of the HV real and synthetic power grids, representative scale-free models of the corresponding power networks have been used. The reliability of the North American eastern and western electric power grids were analyzed based on the Barabási–Albert network model in [28]. In [50], assessment of

vulnerability of power grids is performed through graph theory indexes by using the scale free network models of several IEEE test networks. Scale-free models have also been used to study the vulnerability of the Western United States power grid and Nordic power grid [45], [49]. However, the studies in [25] and [44] have identified that the preferential attachment models are not suitable for modelling the IEEE 300 bus and the Eastern United States power grids.

2.5.2 Fractal-based network models for power networks

Although there are various fractal-based network models developed for different types of real-world networks (e.g. in references [54], [55], [56], [57]), only a few fractal-based models have been found in literature which are specifically developed for power networks.

In [15] and [58], the development of a fractal based networks generating tool for MV and LV power networks of the Great Britain(GB) is explained. Different ranges of fractal dimensions for the distribution of consumer settlements in urban, sub-urban, semi-rural and rural type areas of GB have been used for the development of the tool. The software tool has used an algorithm which simulates the position of consumers in a controllable manner, modelling the characteristics of the positions of real consumer sets, based on different fractal dimensions. Then, realistic networks have been generated by connecting realistic set of consumers' positions with the appropriate branching rates similar to the real networks.

Another fractal graph model for MV and LV distribution grid topologies has been developed in [31]. Two small LV networks and two MV feeders from the distribution networks of Greek power grid have been analysed in order to prove that power networks have fractal like structures. Fractal dimensions of those network samples have been calculated. A two dimensional stochastic dielectric breakdown model (DBM) [60] is utilized to generate virtual distribution networks with different fractal dimensions.

2.5.3 Statistical models/tools for the generation of random-realistic power networks

Models/tools, which are capable of generating large numbers of random-realistic representations of power networks in order to facilitate statistical studies on the power networks are reviewed in this section. The well-known complex networks models and fractal-based network models explained above have provided the foundation for the development of such models/tools in the literature. The related previous research is summarized in Table 2.4.

According to Table 2.4, there are a few statistical models/tools that have been developed for the modelling of HV power networks.

In [20] and [59] the development of a new random topology power grid model, called RT-nested-Smallworld has been presented. Through an investigation of statistical properties of several IEEE networks and two real power grids, the topological and electrical properties of large scale power grids have been identified and quantified. It has been identified that the power grids are sparsely connected with obvious small-world properties and the line impedance has a heavy-tailed distribution. Based on these discoveries, an algorithm has been proposed to generate random topology power grids featuring the same topological and electrical characteristics found from the real data. The motivation behind their study was to find out the classes of network topologies of communication networks that are needed to support the decentralized control of a smart grid through the statistical studies on a large number of power grid test cases.

With similar motivations, a new Cluster and-Connect model is introduced to generate synthetic graphs for HV power networks in [60]. This model has mainly addressed some of the limitations of the above described in the RT-nested-smallworld model.

Table 2.4: A review of the literature on the statistical networks generating tools for power networks.

Work	Voltage level	Name or description of the method/tool	Origin of the supporting network data for the development of the tool	Topological properties of the real networks captured by the model/tool	Electrical properties of the real networks captured by the model/tool	Purpose of the model/tool
[20], [59]	HV	RT-nested-Smallworld model	IEEEE 30, 57, 118, 300 bus test networks, western United States power grid (WSCC) and New York state bulk electricity grid (NYISO)	Node degree distributions and small world properties	Distribution of transmission line impedance	Testing smart grid communication and control networks for the power grids
[60]	HV	Cluster-and-Connect model	Western United States power grid (WSCC) and New York state bulk electricity grid (NYISO)	Clustering, inter- and intra-cluster degree distributions	Distribution of transmission line impedance	Generate synthetic graphs that accurately captures topological and electrical properties
[61]	HV	Random power grid model	HV network in the U.S. state of Florida	Average degree	Impedance distributions	Generate synthetic graphs similar to the real power networks
[15], [58]	MV and LV	Fractal-based statistical tool for design of distribution networks	Distribution networks data of Great Britain from rural, sub-urban and urban areas	Fractal dimension, total network length and branching rates	Consumer demand and network planning standards	Evaluation of alternative design strategies of distribution networks
[31]	MV and LV	Fractal model for distribution grid topologies	2 small LV networks and 2 MV feeders from the Greek power grid	Fractal dimension	Maximum voltage drop, total Power Losses	Test different scenarios and algorithms on MV and LV networks without having to collect and process the real network data

In [61], another generative model for HV power networks has been presented. The model randomly connects N nodes located in a square with a guarantee that there is no isolated node in the network. This model is limited because it only focuses on the averages of degree and impedance distributions. Unlike the previously discussed HV network models, this model does not capture the clustering coefficient and the locally dense clustering behavior of the real power networks.

According to the best of the author's knowledge, fractal-based networks generating tool in [15], [58] and the fractal model in [31] are the only statistical models/tools available in the literature for the modelling of MV and LV networks. The implementation procedures of these fractal models are described in the above section (i.e. Fractal models for power networks).

The fractal-based networks generating tool in [15], [58] has been used to quantitatively assess the impact of alternative distribution network investment plans in terms of capital costs, losses and reliability. This has been done by a statistical assessment approach in which the optimal network design policies are determined by evaluating the costs and benefits of different designs applied on a number of networks with similar properties.

The fractal model developed in [31] has been validated only on a limited number of real networks. More networks need to be analyzed for a better classification of MV and LV networks and in order to make a robust validation of this fractal model.

2.6 Summary and research gaps

- According to the review of the literature on impact assessment studies of LCTs, most of the previous research are based on one or few real or synthetic network samples. Although, such studies give an insight into the likely problems that may be experienced in the distribution networks, conclusions made from those studies have limited applicability to other networks.
- Researchers have identified the need of statistical studies carried out on large numbers of real/realistic networks to support robust decision making on the impacts of LCTs in distribution networks. The main obstacles for researchers to conduct such statistical studies have been identified; limited access to real-world network data, limited availability of representative network models and limited availability of open source tools for the generation of representative test cases of the real networks.
- The need of analyzing the statistical properties of the power networks in order to develop statistical models which can generate random-realistic power networks has been identified. A review of the literature on investigations of the statistical properties of power networks was presented. Complex networks analysis techniques together with graph theory applications have been widely used in the previous research to analyze the topological and electrical properties of the real and synthetic power networks. In addition, fractal properties of the power networks have also been analyzed. Majority of the related research has been carried out for the HV power networks and are supported by a limited number of real or synthetic networks.
- Statistical models that are available in the literature for the modelling of power networks were also reviewed. Well-known complex networks models (i.e. random graph, small-world and scale-free network models) have been widely used to model the

HV power networks. The main goal of such studies was to find out the relation between the vulnerability and topological and electrical structures of the power networks. A few, (complex networks theory and fractal theory-based) new statistical models/tools that have been specifically developed for the modelling of power networks were found in the literature. Majority of such models/tools have been developed for HV power networks. Statistical models/tools that has been developed for the MV and LV networks are based on the fractal generating methods. So far, these fractal models are limited for certain geographic areas where the fractal properties of the real-world distribution networks have been observed.

Chapter 3: Topological properties of medium voltage electricity distribution networks

3.1 Introduction

Characterising the important statistical properties of different types of real world electricity distribution networks are required to generate random, realistic test network models for various simulation studies. In this thesis, the statistical properties of electricity distribution networks are studied under two categories, namely topological and electrical properties. An investigation of the topological properties of real-world, MV electricity distribution networks is presented in this chapter.

Topological properties of a distribution network describe how different network components are located and connected to each other, which is critical to evaluate the network performance. Simple models which assume equal spacing between the consumers and equally spaced substations are not accurate enough to represent the topological features of real distribution networks.

Both topological and electrical properties of distribution networks are decided by many factors such as, spatial distribution of the consumer load, network planning guidelines and recommendations, social and financial factors. Due to the effect of these factors, each electricity distribution networks may have observable differences in their topological and electrical properties. However, regardless of the detailed differences of these topological and electrical properties, similar types (e.g. rural/urban) of distribution networks share some common statistical characteristics. For example, in rural networks consumer locations tend to be

distributed in a more clustered fashion with large open areas dedicated to farms and green spaces, while in urban networks consumer locations are usually evenly distributed.

To capture these features, a rigorous analysis of real-world networks is required. From the literature review in the Chapter 2, it has been identified that most of the topological investigations are carried out on the HV level on the power networks. Topological structures of HV grids are different from that of MV and LV grids. The HV transmission and sub-transmission is usually a meshed system, but distribution networks (MV and LV) are mainly radial structures. The research findings in the HV network analysis cannot be directly used in MV and LV distribution networks. Also, majority of the previous studies were supported by a limited set of real network data or only a limited set of topological properties were analysed. Hence a wide ranging topological analysis supported by a large amount of real world network data is required for electricity distribution networks.

The motivation behind the topological investigation presented in this chapter comes from finding out answers to the following two questions;

i. What are the key topological properties that characterize the realistic nature of different types of electricity distribution networks?

ii. Is it possible to efficiently generate ensembles of random but realistic network topologies similar to real electricity distribution networks in order to conduct a large amount of simulation studies?

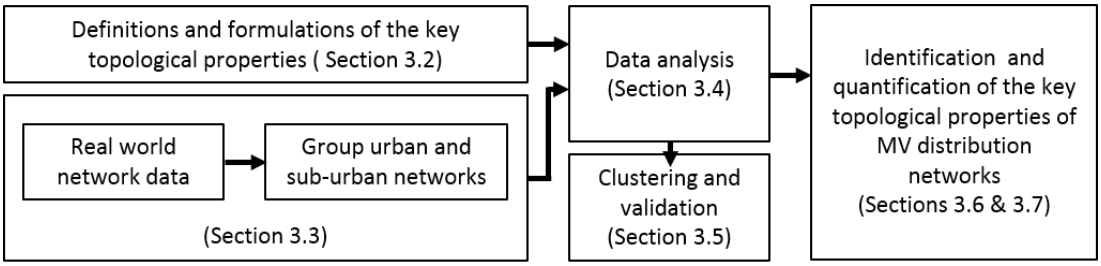


Figure 3.1: Schematic overview of the topological study.

The schematic overview of the study of topological properties conducted in this chapter, is shown in Figure 3.1. First, some of the essential definitions and formulations of the key properties used in the topological investigation of the distribution networks are described. Then, a thorough investigation of topological properties of real distribution networks at MV level is conducted using real-world data. The sub-urban and urban grouping of the distribution networks have been validated using a clustering approach. Finally, the key topological properties that characterize different types (urban and sub-urban) of distribution networks have been identified and quantified.

3.2 Definitions and formulations of the key topological properties

In order to understand and model a topology of a real-world complex network such as an electrical power grid, a right set of tools and techniques are required. As seen in the literature review in Chapter 2, most widely used techniques and tools are coming from the fields of Complex Network Analysis (CNA) and graph theory [33]. Ignoring the 3-dimensional features such as very tall buildings and elevation of the equipment, the electrical power grid can be considered as a 2-dimensional grid composed of various elements such as transmission and distribution lines, distribution transformers and switchgears. A graph model can be easily constructed by taking into consideration the topological relationship between these elements.

This section describes some of the essential definitions and formulations of the key properties used in the topological investigation of the distribution networks. The physical interpretation of

graph related properties with reference to the electrical power networks are also described in this section. Figure 3.2 summarizes the fundamental approach and the key topological features used in this study.

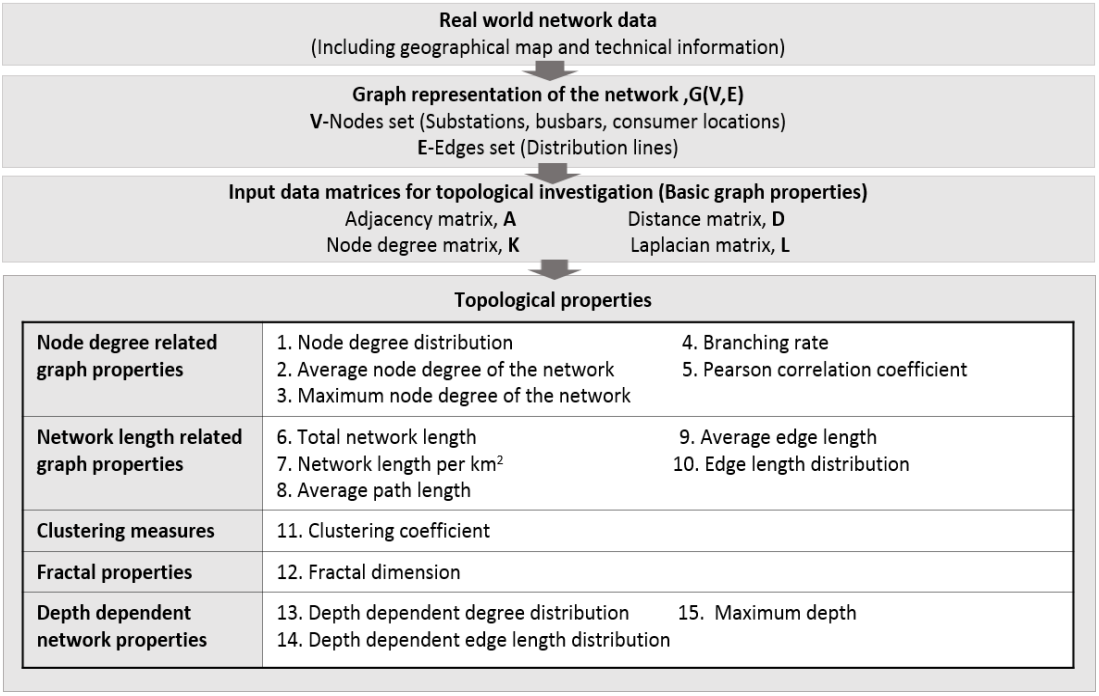


Figure 3.2: Feature extraction.

3.2.1 Basic graph properties

A graph $\boldsymbol{G}$ consists of a collection of vertices $\boldsymbol{V}$ (nodes) and a collection of edges $\boldsymbol{E}$: $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$. With reference to an electricity distribution network, $\boldsymbol{V}$-nodes set include substations, distribution transformers, switches, busbars and consumer locations. Edges set $\boldsymbol{E}$, stands for the physical connections between the nodes through underground (UG) cables and overhead (OH) distribution line segments. Since the collected real network data includes the actual geographical location of the electrical components, the information about nodes are extracted with the x, y coordinates. The edges are represented using the start and end x, y coordinates of the nodes. For an electrical power network with N nodes and M edges,

$$\boldsymbol{V} = [x, y]_{N\times 2} \tag{3.1}$$

$$\boldsymbol{E} = \left[start_x, stary_y, end_x, end_y, circuit\_id\right]_{M\times 5} \tag{3.2}$$

In real networks, very often two or more circuits share the same towers. In that case, to distinguish them a fifth dimension called $circuit\_id$ is added to $\boldsymbol{E}$.

- **Adjacency matrix ($\boldsymbol{A}$)**

Connectivity between nodes in the graph is represented by an Adjacency Matrix '$\boldsymbol{A}$'. As the electrical power grid can be considered as an undirected graph, the adjacency matrix of a power grid becomes a symmetric $N \times N$ matrix. The element $\boldsymbol{A}_{ij}$ becomes 1 if there exists a link between nodes i and j, otherwise $\boldsymbol{A}_{ij}$ equals to 0 (3.3).

$$\boldsymbol{A}_{ij} = \begin{cases} 1, & if (i,j) \in \boldsymbol{E} \\ 0, & if\ (i,j) \notin \boldsymbol{E} \end{cases} \tag{3.3}$$

- **Distance matrix ($\boldsymbol{D}$)**

The distance matrix '$\boldsymbol{D}$' is defined for a graph using the edge lengths between nodes. The element $\boldsymbol{D}_{ij}$ becomes the length of the edge e_{ij} if there exists an edge between nodes i and j. Otherwise $\boldsymbol{D}_{ij}$ is equal to zero.

$$\boldsymbol{D}_{ij} = \begin{cases} \sqrt{\{\boldsymbol{V}(i,1) - \boldsymbol{V}(j,1)\}^2 \ + \{\boldsymbol{V}(i,2) - \boldsymbol{V}(j,2)\}^2}\ , & if (i,j) \in \boldsymbol{E} \\ 0, & if\ (i,j) \notin \boldsymbol{E} \end{cases} \tag{3.4}$$

- **Degree matrix ($\boldsymbol{K}$)**

The degree of a node i which is denoted as k_i, is the number of edges incident to that node and is obtained using the adjacency matrix. The values obtained for the node degrees are used to construct the Degree Matrix '$\boldsymbol{K}$' which is an $N \times N$ diagonal matrix.

$$k_i = \sum_{j=1}^{N} \boldsymbol{A}_{ij} \tag{3.5}$$

$$\boldsymbol{K}_{ij} = \begin{cases} k_i\ , & if\ i = j \\ 0\ , & otherwise \end{cases} \tag{3.6}$$

- **Laplacian matrix ($\boldsymbol{L}$)**

Laplacian matrix '$\boldsymbol{L}$' is also useful in obtaining graph properties of power networks which is expressed as,

$$\boldsymbol{L} = \boldsymbol{K} - \boldsymbol{A} \tag{3.7}$$

A connected component in graph theory refers to a set of vertices in a graph that are linked to each other. For instance, a radial 10kV network supplied by one 35kV/10kV substation can be considered as one connected component (ignoring the connections to the main grid from the 35kV side of the main supply point). According to this definition, the number of connected components in a radial electricity distribution network of a certain voltage level (in a given area) is equal to the number of main grid supply points in the network. From the graph theory definitions, the number of connected components in a graph is equal to the number of times 0 appears as an eigenvalue in the Laplacian matrix of the graph.

Most of the graph-related properties of the networks described below are derived using the above basic graph properties $\boldsymbol{A}$, $\boldsymbol{D}$, $\boldsymbol{K}$ and $\boldsymbol{L}$.

3.2.2 Node degree related graph properties

Node degrees can be used to identify the key components in a network. In power networks, nodes such as substations have a high node degree compared to the other nodes. Also, degrees, and notably degree distributions can be used to derive information on the structure of a

network. For example, if most node degrees are the same the network is more or less a regular network in which vertices have equal roles.

- **Nodes degree distribution**

The nodes degree distribution $P(k)$ of a network is defined as the fraction of nodes in the network with degree, k. If the total number of nodes in the network is N and n_k of them have degree k, $P(k)$ is defined in Equation (3.8).

$$P(k) = \frac{n_k}{N} \tag{3.8}$$

A power law relationship between $P(k)$ and k is as shown in Equation (3.9) where, γ is a parameter whose value is typically ≥ 1.

$$P(k) \propto k^{-\gamma} \tag{3.9}$$

As seen in the literature review in Chapter 2, in vulnerability assessment studies on HV transmission grids, this power law relationship between $P(k)$ and k has been broadly analysed [28], [34]. However, for weakly meshed and radial networks in distribution level a power law distribution for $P(k)$ cannot be observed. But still it is worth to observe the behaviour of $P(k)$ for similar types of distribution level networks.

- **Average node degree**

The average node degree k_{avg}, of a graph $\boldsymbol{G}$ is also an important measure about the structure of the network. If k_{avg} > 2 the network has a meshed structure.

$$k_{avg} = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2M}{N} \tag{3.10}$$

- **Branching rate**

Another node degree related measure is the branching rate b_r, which gives an indication of how much a given network tends to branch out. For instance, the urban distribution networks tend to branch out more compared to the rural distribution networks. Branching rates are usually different for different types of networks (rural, urban), for different voltage levels, and for different locations, i.e. close to the supply points or close to the customer points.

$$b_r = \frac{\text{Number of nodes with degree} \geq 3}{\text{Total number of nodes in the network}} \tag{3.11}$$

- **Pearson correlation coefficient**

A network is said to show assortative mixing if there exists a correlation between nodes of similar degree. Assortativity property of the networks is examined in terms of node degrees using the Pearson correlation coefficient ρ:

$$\rho = \frac{M^{-1}\sum_i j_i k_i - [M^{-1}\sum_i \frac{1}{2}(j_i + k_i)]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - [M^{-1}\sum_i \frac{1}{2}(j_i + k_i)]^2} \tag{3.12}$$

where j_i, k_i are the degrees of the vertices at the ends of the i^{th} edge, with i = 1, ..., M. A positive value for ρ indicates correlation between nodes with similar degree and negative values for ρ indicates the relationship between nodes of different degree which is called disassortative mixing [62].

3.2.3 Network length related graph properties

In power system terminology, network length of an MV distribution network refers to the total length of the OH and UG electricity distribution line segments. The network length is a critical parameter for electrical power networks since it impacts on a number of technical and economic factors such as the voltage drops, power losses and the cost of UG cables/OH lines in the

network. It is also important when describing the network topology realistically. An edge can be referred to a feeder section in an electrical distribution network. Some of the length related network properties are defined below.

- **Total network length**

Total network length L_{total}, is the addition of all the edge lengths in the network.

- **Average edge length**

Average edge length e_{avg} is obtained by dividing the total network length by the total number of edges.

- **Average path length**

A path is a continuous sequence of edges from one vertex to another. Length of a path is the addition of the lengths of all the edges in the path. The 'geodesic distance' between nodes v_1 and v_2, denoted as $l(v_1, v_2)$ is the length of the shortest path between v_1 and v_2. Diameter of a network is defined as the longest graph geodesic between any two graph vertices v_1, v_2 of a graph. Then the average path length l_{avg} is defined as the average length along the shortest paths for all possible pairs of network nodes.

$$l_{avg} = \frac{1}{N(N-1)} \sum_{i \neq j} l(v_i, v_j) \tag{3.13}$$

- **Edge length distribution**

Similar to the degree distribution, the edge length distribution $P(e)$ of a network is defined as the fraction of edges in the network, with length e. Here, length e represents a range of edge

lengths. If the total number of edges in the network is M and m_e of them fall in to the length range e, $P(e)$ is expressed as:

$$P(e) = \frac{m_e}{M} \tag{3.14}$$

Some network models assume equal spacing between nodes. However, this is not true for most real networks. Therefore, the edge length distribution is an important feature to be considered in the network modelling.

3.2.4 Clustering of nodes

- **Clustering coefficient**

In graph theory, clustering coefficient C is a measure of the degree to which nodes in a graph tend to cluster together. The clustering coefficient is defined as the average of the clustering coefficient for each node C_i.

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{3.15}$$

C_i is expressed as the ratio of number of edges between the neighbours of node i, $\lambda_G(i)$ to the total number of edges that can exist among neighbours of node i, $\tau_G(i)$ [52].

$$C_i = \frac{\lambda_G(i)}{\tau_G(i)} \tag{3.16}$$

The clustering coefficient of a radial network is zero. Therefore, in order to analyse the MV and LV radial networks this measure is not very useful. But in HV network studies this measure has been widely used (Section 2.4, Chapter 2).

3.2.5 Fractal properties

As explained in Chapter 2, a few researchers have observed that the real-world electricity distribution networks also consist of self-repeating patterns across all scales similar to the fractals [15], [63].

- **Fractal dimension**

Fractal dimension is an important property of fractals which provides a statistical index for the complexity of the fractal[36]. If the fractal dimension of an object needs to be explained, the box counting method provides the practical solution for that. In this method, the fractal object (e.g. image of the distribution network layout) is covered with boxes with equal sides ε, and find how the number of boxes $n(\varepsilon)$ which include the fractal object changes with the box size. A network is said to be fractal if the box counting dimension exists for that network. The box-counting dimension D_f is defined as [64]:

$$D_f = lim_{\varepsilon \to 0} \frac{\log n(\varepsilon)}{\log(\frac{1}{\varepsilon})} \tag{3.17}$$

3.2.6 Depth property and depth dependent network properties

- **Depth of a node**

The depth d, of a node x, in a graph is defined as the number of edges n_e, from the root node r, to the node x. The depth of a node from the given root node can be obtained using Dijkstra shortest path algorithm [65]. In this study, the HV/MV substation is considered as the root of a MV radial network. A simple connected graph with no loops is called a tree. The electricity distribution networks at MV and LV levels predominantly have tree like (radial) structures.

$$d_{r,x} = n_e(r, x) \tag{3.18}$$

- **Maximum depth of the network**

Maximum depth d_{max} is the number of edges along the longest path from the root node down to the farthest leaf node.

During this study, analysing the depth dependent network properties was identified as an effective approach in both topological feature identification and the network model development. More details can be found in the section 3.7.

3.3 Real-world power grid data for topological investigation

The topological investigation of the real-world networks was conducted using network data collected from China, covering urban and sub-urban areas. These data include the detailed technical and geographical information of transmission, sub-transmission, and distribution level networks and the population data of the supplied areas of the networks. Since the present work is mainly focused on MV level networks, only the 10kV voltage level network information was used in this study. 10kV network is the most extensive forms of the main distribution system in both rural and urban areas in China. It is supplied by medium to large primary substations, where power is transformed from 35kV/10kV. As most of the MV distribution networks are operated in radial mode in the normal operating conditions [66], radial structures of the networks are considered in this study.

A graph representation of each network was obtained. A graph representation obtained from the geographical layout of a single network, is shown in Figure 3.3(a). The red nodes represent the 10kV level consumers, 10kV/400V distribution transformers (secondary substations) and busbars. The green circle represents the main grid supply point (primary substation) of the 10kV network.

Table 3.1 summarizes the basic information available for the selected 30 networks at the 10kV level. The networks were categorized as sub-urban and urban depending on the population density. According to the Demographic Yearbook 2013 published by United Nations, the definition of 'urban' for cities in china is defined as the areas with population density higher than 1500 people per square kilometre [67]. The distribution of the population density of the networks under study is shown in Figure 3.3(b).

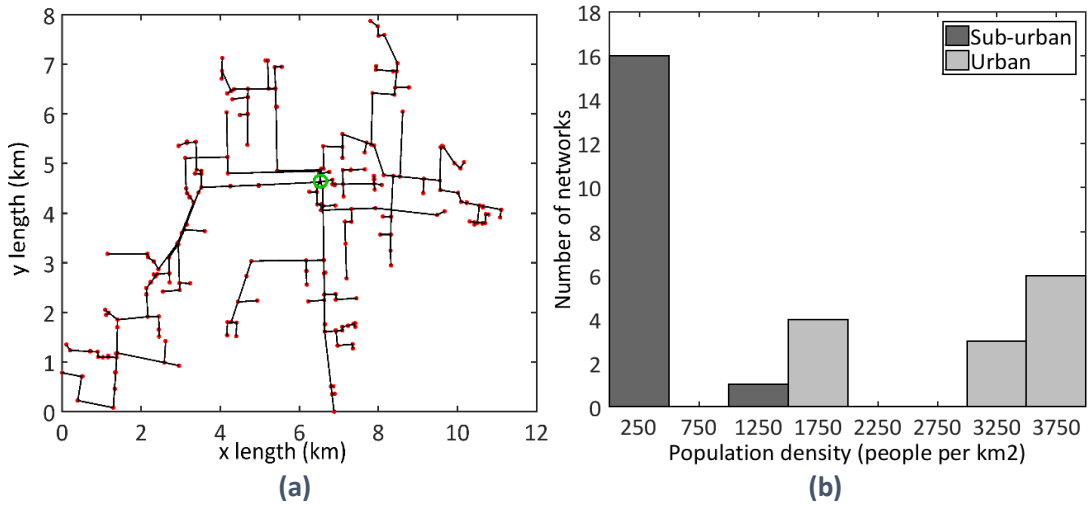


Figure 3.3: (a) Graph representation of network no. 1 (b) Distribution of population densities of the networks.

According to Figure 3.3(b) out of 30, 10 kV networks, 17 networks are in the sub-urban category while the other 13 networks fall in to the urban category.

Table 3.1: Basic network information of 30 networks at 10kV level.

Network ID	Area (km^2)	Population Density ($/km^2$)	Total network length (km)	Number of nodes	Number of Edges	Network ID	Area (km^2)	Population Density ($/km^2$)	Total network length (km)	Number of nodes	Number of Edges
1	66.2	482	76.4	254	253	**16**	101	256	103	377	376
2	142	328	88.5	399	398	**17**	44.3	1054	84.1	384	383
3	64.5	405	64.2	285	284	**18**	9.8	1939	28.3	234	233
4	86.2	377	80.6	350	349	**19**	9.2	1935	29.6	237	236
5	33.8	376	35.5	136	135	**20**	10	1930	34.2	331	330
6	108.5	256	78.8	204	203	**21**	8	1750	34.2	205	204
7	86.2	229	66.8	226	225	**22**	14	3200	38	328	327
8	93.8	228	81.5	246	245	**23**	16	3200	52.5	400	399
9	69	464	90.6	308	307	**24**	14	3200	31.2	267	266
10	44.3	361	32.1	100	99	**25**	7	3600	28.7	227	226
11	78.5	369	61.1	196	195	**26**	7.5	3600	29	217	216
12	84.1	473	87.3	321	320	**27**	7	3600	17.5	115	114
13	69.1	411	71.6	220	219	**28**	7.5	3600	14.7	114	113
14	80.9	449	54.2	173	172	**29**	7.5	3600	16.1	153	152
15	125.4	364	95.6	351	350	**30**	8	3600	30.4	265	264

3.4 Quantification of the topological properties

Table 3.2 presents the topological properties of the 30 chosen networks at 10kV level.

In order to compare the results shown in Table 3.2, probability distributions of the topological properties of both sub-urban and urban networks were obtained. Figure 3.4 shows the comparative probability distribution plots for the two types of networks, arranged back to back on the x-axis (probability of occurrence). For one topological property, the same bin size and the same number of bins were used to generate the probability distributions for both types of networks.

It can be observed from Figure 3.4 (sub-graphs with letter A) that, some of the topological properties such as, nodes per km^2, network length per km^2, average edge length and average path length are able to clearly characterise the topological differences of the two types of networks. The ranges of the variation of the above four properties are summarised in Table 3.3.

The probability distributions of the branching rate, maximum node degree, fractal dimension and maximum depth have some noticeable differences (i.e. probability distribution plots of the two network types are biased into different directions) of the two types of networks. However, still there are some overlapping of the values (sub-graphs with letter B). According to Figure 3.4 (sub-graphs with letter C), properties such as Average node degree and Pearson correlation coefficient do not give clear information to characterise urban and sub-urban networks.

Table 3.2: Graph related properties of the 30 networks at 10 kV level.

Network ID	Number of Nodes per 1km^2	Network length p 1km^2 (km)	Average node degree	Branching rate	Maximum Node degree	Pearson correlation coefficient	Average edge length (km)	Average path length (km)	Fractal dimension	Max Depth
1	4	1.154	1.992	0.324	7	0.534	0.302	7.408	1.301	28
2	3	0.623	1.995	0.313	5	0.559	0.222	10.356	1.270	54
3	4	0.995	1.993	0.365	4	0.457	0.226	7.551	1.275	35
4	4	0.934	1.994	0.377	7	0.516	0.231	6.445	1.331	28
5	4	1.052	1.985	0.346	4	0.421	0.263	5.404	1.266	23
6	2	0.726	1.990	0.363	5	0.504	0.388	8.418	1.261	26
7	3	0.775	1.991	0.292	3	0.776	0.297	8.518	1.273	34
8	3	0.868	1.992	0.350	4	0.356	0.332	9.694	1.276	35
9	4	1.316	1.994	0.361	5	0.477	0.295	8.179	1.323	38
10	2	0.724	1.980	0.330	3	0.394	0.324	4.546	1.228	25
11	2	0.779	1.990	0.352	4	0.597	0.313	7.625	1.270	24
12	4	1.038	1.994	0.336	6	0.616	0.273	7.983	1.319	35
13	3	1.036	1.991	0.341	3	0.600	0.327	9.840	1.285	38
14	2	0.670	1.988	0.358	4	0.353	0.315	6.513	1.266	27
15	3	0.762	1.994	0.288	7	0.475	0.273	9.792	1.296	51

16	4	1.020	1.995	0.358	8	0.612	0.274	9.452	1.319	37
17	9	1.899	1.995	0.299	9	0.363	0.220	6.820	1.344	36
18	24	2.885	1.991	0.350	7	0.538	0.121	2.879	1.304	23
19	26	3.222	1.992	0.371	7	0.682	0.126	3.388	1.302	38
20	33	3.418	1.994	0.390	7	0.606	0.104	2.641	1.336	30
21	26	4.275	1.990	0.263	9	0.502	0.168	3.738	1.281	22
22	23	2.711	1.994	0.348	8	0.499	0.116	3.788	1.324	45
23	25	3.281	1.995	0.370	8	0.678	0.132	4.370	1.341	29
24	19	2.230	1.993	0.397	9	0.101	0.117	3.415	1.352	33
25	30	3.828	1.991	0.370	6	0.633	0.127	4.494	1.353	40
26	29	3.867	1.991	0.373	6	0.580	0.134	3.578	1.341	29
27	15	2.336	1.983	0.365	7	0.390	0.154	2.583	1.327	20
28	15	1.960	1.982	0.272	6	0.491	0.130	2.232	1.319	20
29	20	2.152	1.987	0.366	7	0.542	0.106	2.053	1.369	31
30	35	4.048	1.992	0.343	10	0.226	0.115	2.507	1.344	21

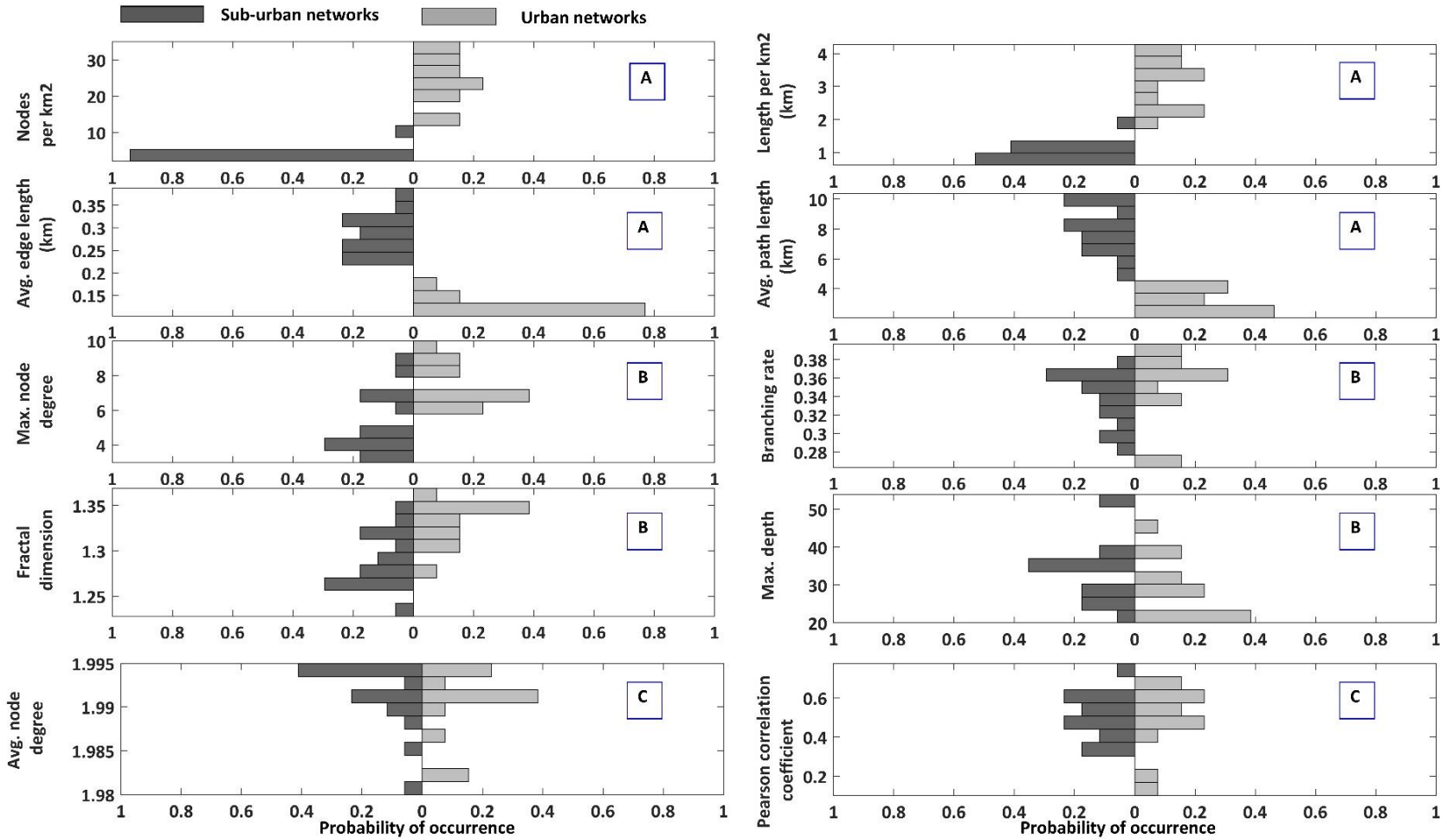


Figure 3.4: Comparison of the topological properties of urban and sub-urban networks.

Table 3.3: The ranges of variations of the topological properties of sub-urban and urban distribution networks.

Topological property	The range of variation	
	Sub-urban networks	Urban networks
Nodes per km2	2 - 9	15 – 35
Network length per km	0.6 km - 2.0 km	2.0 km – 4.2 km
Average edge length	0.2 km – 0.4 km	0.1 km – 0.2 km
Average path length	4.5 km – 10.5 km	2.0 km – 4.5 km

3.5 Validation of the networks classification using topological properties

A validation of the network classification with the population density parameter (i.e. as urban and sub-urban networks), is validated in this section using a clustering approach.

- **K-means clustering**

Clustering in general, is defined as the grouping of similar objects. The *k*-means clustering algorithm was used to classify the data set in Table 3.2 into a set of clusters. The *k*-means clustering algorithm aims to partition a number of observations into *k* number of clusters in which each observation belongs to the cluster with the nearest mean. Each observation is a *d*-dimensional real vector. Euclidian distances are used to calculate the distance from the observation to the mean.

The implementation of the *k*-means algorithm is as follows; (i) make initial guesses for the means m_1, m_2,.. m_k. (ii) use the estimated means to classify the samples into clusters. (iii) for *i* from 1 to *k*, replace m_i with the mean of all of the samples for cluster *i* (iv) repeat steps 2 and 3 until there are no changes in any mean [68], [69].

Table 3.2

Network	Property					
	p_1	p_2	p_3	..	....	p_{10}
1						
2						
3						
: :						
30						

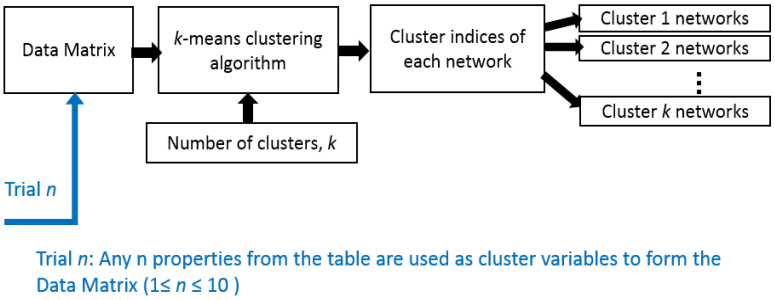


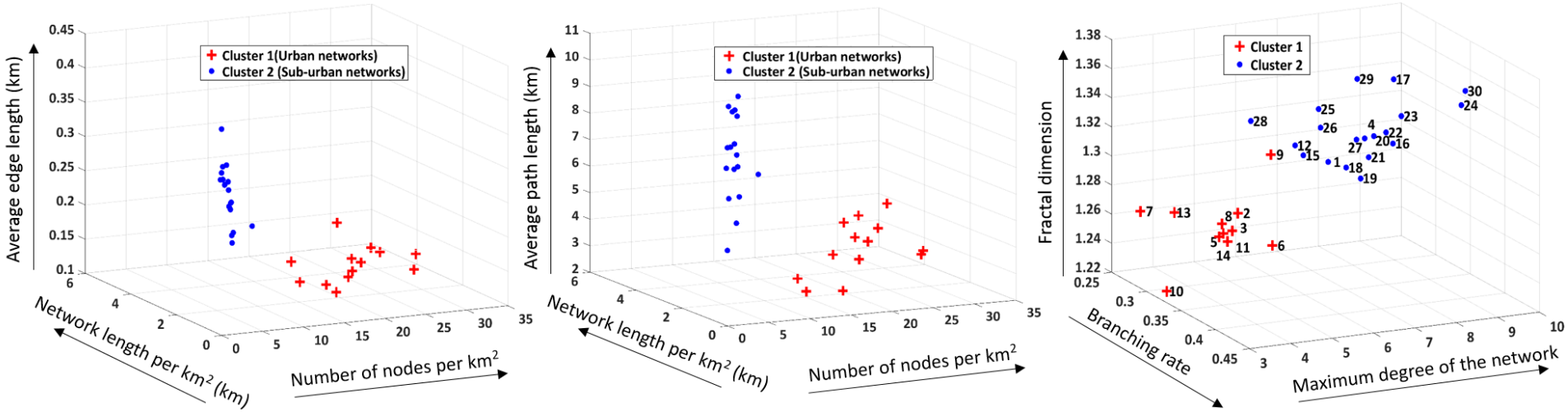
Figure 3.5: The k-means cluster analysis.

Figure 3.5 illustrates the procedure of the *k*-means cluster analysis used in this study. According to the above explanation of *k*-means clustering, the 30 networks used in the study represent 30 observations, and each observation is a 10-dimensional real vector. The 10 dimensions (cluster variables) are the 10 topological properties listed in Table 3.2. The input data matrix for the *k*-means algorithm was formed using different subsets of the properties from Table 3.2 to find out which subsets of the parameters together can effectively characterize the two network types (urban and sub-urban). Therefore, it was assumed that the number of clusters *k* is known for the data set (*k*=2). Two clusters are the urban and sub-urban networks. Then, the *k*-means algorithm was used to group the 30 networks into two clusters.

Trial *n*, refers to the exercise where any n properties from Table 3.2 are used as cluster variables to form the Data Matrix in Figure 3.5. Trial 10 and trial 3 were used as examples for the discussion of the clustering results in this section. In the trial 10 all the 10 topological parameters in Table 3.2 were used as cluster variables. The grouping done by the clustering in trial 10 exactly followed the sub-urban and urban classification done by the population density parameter. Hence trial 10 was used to validate the urban, sub-urban classification of the networks.

The results shown in Figure 3.6 are graphical representations of the case where, different subsets of 3 topological parameters were chosen as the cluster variables (trial 3). In the first two cases (Figure 3.6 (a) and (b)), the selected sets of cluster variables were able to group the

network sample into two clusters accurately, as defined by the population density of the networks. However, the third set of cluster variables shown in Figure 3.6(c) did not cluster the network sample into the right groups. Some of the sub-urban type networks were fallen into the urban category and also the data points in each cluster seemed to be much more dispersed that the previous two cases. This observation explains the importance of feature selection when characterizing different network types.



(a) (b) (c)

Figure 3.6: Cluster assignments.

3.6 Identification of key topological properties

From the results so far, it is evident that the node degree and edge length related topological properties are key to characterise different types of networks (most of the above listed topological properties are related with node degrees and edge lengths).

In Section 3.4 all the node degree and edge length related topological properties are evaluated using a single value to describe each network. However, by investigating the probability distributions of these topological properties more details about the network topologies can be extracted.

- **Probability distributions of key topological properties**

Figure 3.7 (a) and (b) show the edge length distributions of sub-urban and urban networks respectively. Light blue and light red curves in both figures represent the edge length distribution of a single network. The dashed dark blue and dark red lines show the average curves of the edge length distributions of all the networks in the corresponding figure. The average curve was obtained by taking into consideration the edge lengths in all the networks of one type as one set and by getting the probability of occurrences for the whole set.

The curves of average edge length distribution of sub-urban and urban types of networks are compared in Figure 3.7(c). It was observed that edge length distributions of both types of networks follow negative exponential patterns. The edge length distribution of urban networks has a faster decay compared to the sub-urban networks and this observation explains that the urban networks have a considerably higher fraction of shorter edge lengths compared to the sub-urban networks.

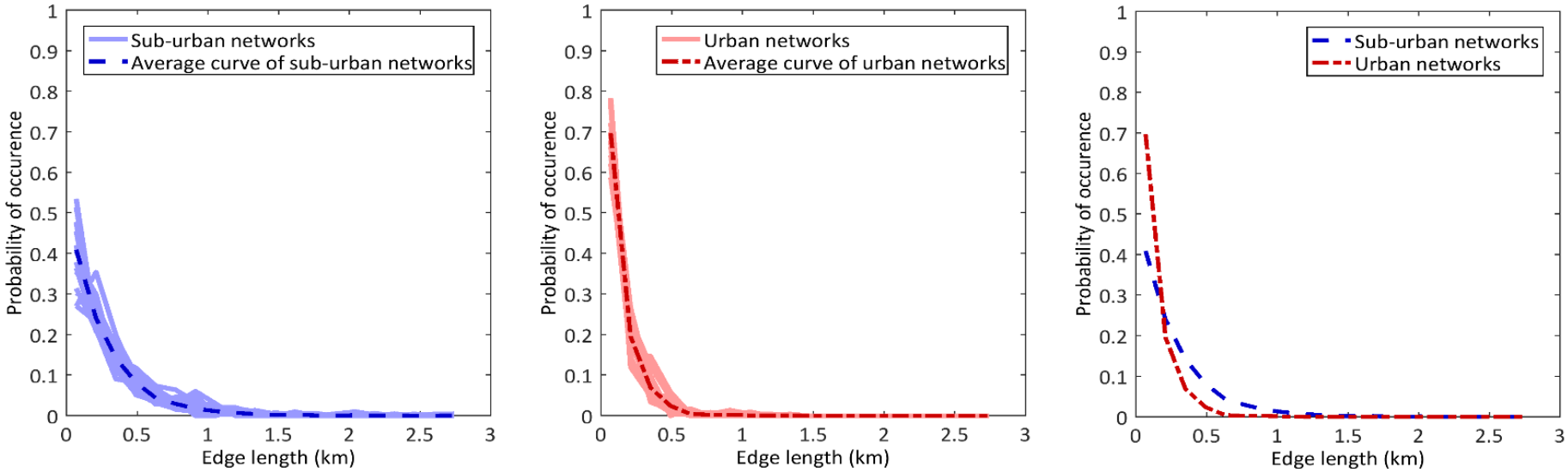


Figure 3.7: Edge length distributions of; (a) sub-urban networks, (b) urban networks and (c) comparison of edge length distributions of sub-urban and urban networks.

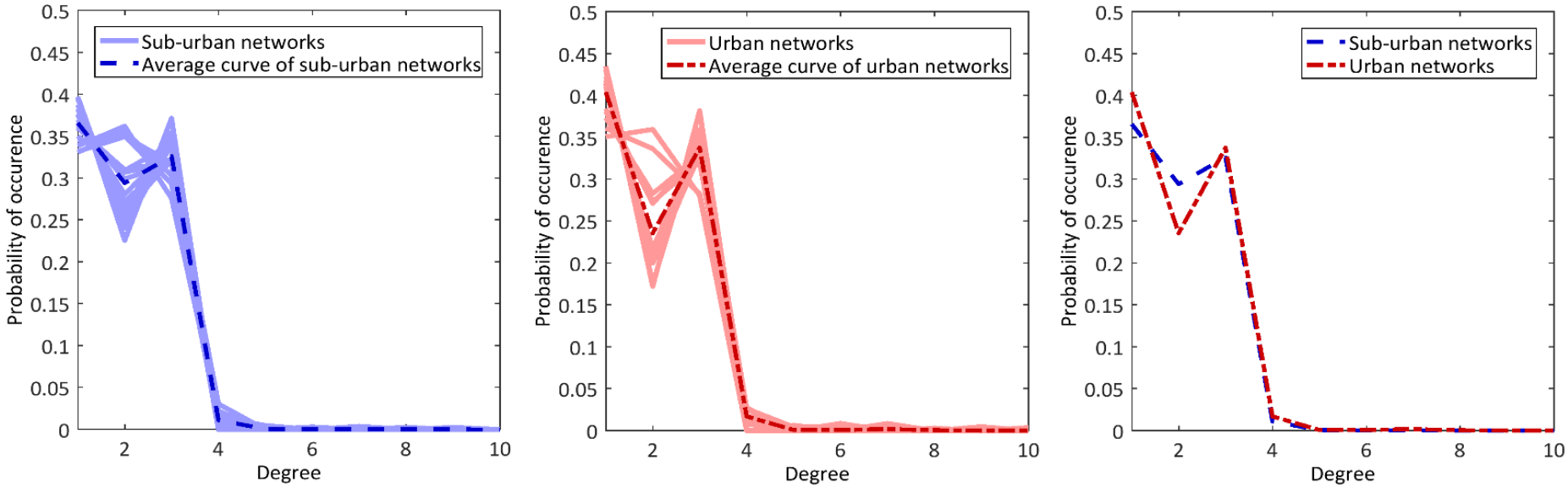


Figure 3.8: Node degree distributions of; (a) sub-urban networks, (b) urban networks and (c) comparison of node degree distributions of sub-urban and urban networks.

Similarly, degree distributions of the sub-urban and urban networks are shown in Figure 3.8(a) and (b) respectively. The average degree distribution curves for the two types of the networks are compared in Figure 3.8(c). The curves do not follow any well-known distribution. However, it is noticeable that the number of nodes with a degree of 2 in most of the urban networks is less than the sub-urban networks. This implies that the urban networks tend to have more branches (nodes with degree ≥3) and leaf nodes (nodes with degree 1) than sub-urban networks. Also, the maximum degree observed in sub-urban networks is 7 while maximum degree of the urban networks is up to 10.

3.7 The depth dependent topological properties

Electrical power grid is an evolving network, with new nodes (substations and consumers), and edges (power distribution lines) added with time. Similar to most of the real-world networks, the development and evolution of the electrical power networks is closely related to the factors such as geographical environment, population distribution, social and economic development. Due to these factors, different networks have observable topological differences.

It was observed that, the electrical power networks used in this study have considerably different graph related properties at different depths of the networks. For example, in urban networks consumer locations are evenly distributed with compared to the sub-urban networks. Therefore, urban networks has a higher network density compared to the sub-urban networks.

3.7.1 Definition of 'depth levels' for the analysis of topological properties

In this study, topological properties of the networks are observed depending on the depth of the nodes as; closer to the supply point (level 1), at the middle level (level 2) and at the furthest away area from the supply point (level 3) respectively.

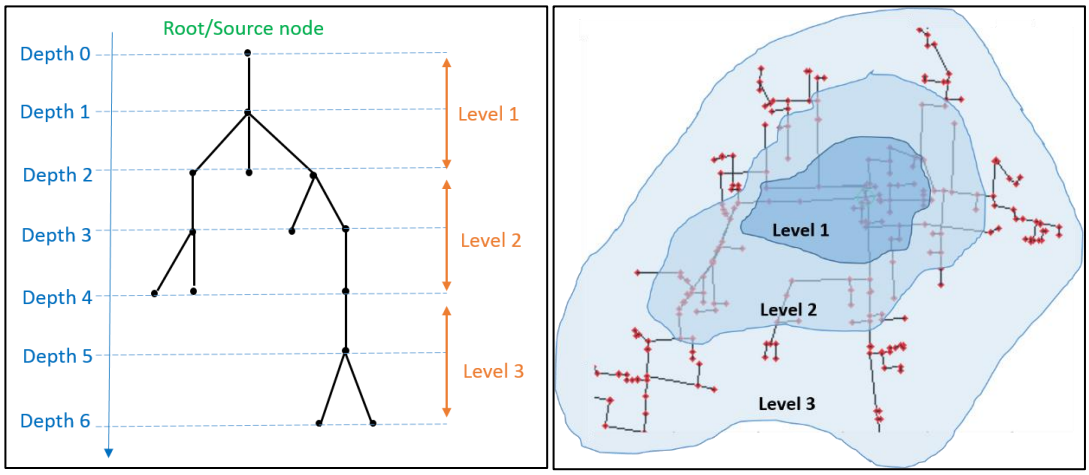


Figure 3.9: (a) The concept of depth of a node (b) The idea of the levels of a network along the depth.

The network is divided into several levels along the depths as shown in Figure 3.9(a). In this study, the network is divided into three levels l_1, l_1 and l_1 along the depth. The number of levels chosen for the depth dependent study, can be varied. A sensitivity study with different number of levels can be conducted to identify the best number of levels to capture the depth dependent properties of different types of networks. However, this part of the study is not covered in this thesis. Figure 3.9(b) illustrates segmenting a radial network into levels.

$$l = \begin{cases} l_1; & 0 < d \leq \dfrac{d_{max}}{3} \\ l_2; & \dfrac{d_{max}}{3} < d \leq \dfrac{2d_{max}}{3} \\ l_3; & \dfrac{2d_{max}}{3} < d \leq d_{max} \end{cases} \tag{3.19}$$

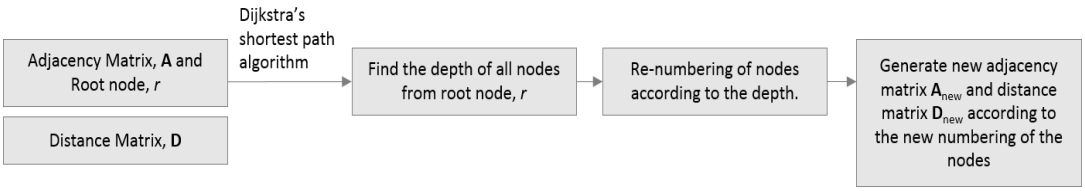


Figure 3.10: Algorithm to obtain depth dependent network properties.

In order to obtain the depth dependent properties for a tree like graph, the adjacency '$\boldsymbol{A}$' and distance '$\boldsymbol{D}$' matrices can be re-organized as '$\boldsymbol{A}_{new}$' and '$\boldsymbol{D}_{new}$' following the new node

identifiers given according to the depth of the node (Figure 3.10). For example, the root/source node is now numbered as 'node 1' and the nodes immediately connected to the root node takes the next consecutive numbers for their node identifiers. Submatrices of the $\boldsymbol{A}_{new}$ and $\boldsymbol{D}_{new}$ are used to derive the depth dependent degree and edge length distributions of the network. For example, the edge length distribution of the network at level 1, (l_1) is obtained from the values in submatrix $\boldsymbol{D}_{l1}$ of $\boldsymbol{D}_{new}$, using the basic definition in the Equation (3.14). Similarly, the values in submatrices $\boldsymbol{D}_{l1}$ and $\boldsymbol{D}_{l1}$ are used to obtain the edge length distributions of level 2 and level 3 respectively. The idea is illustrated in the Figure 3.11.

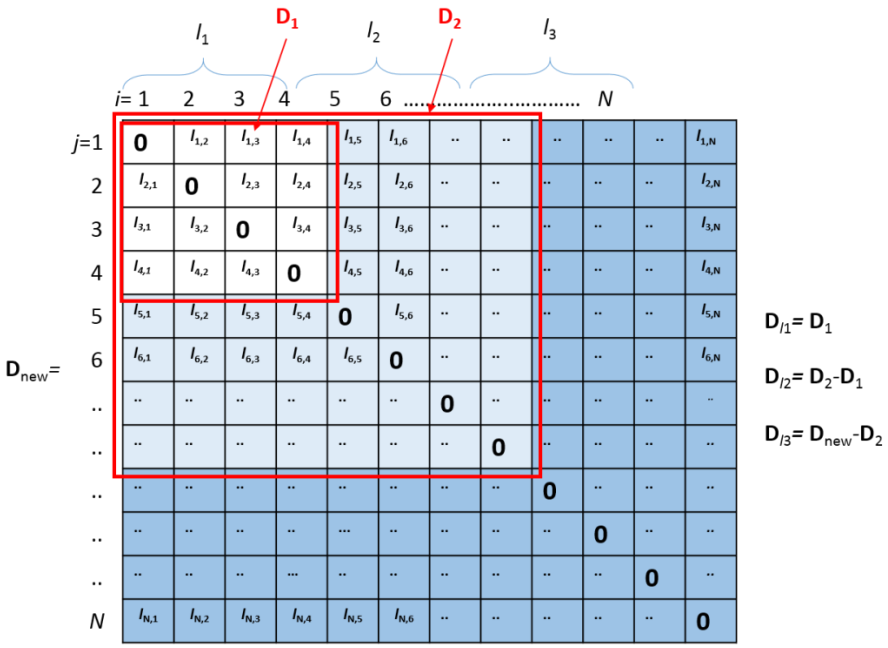


Figure 3.11: New distance matrix (node numbers 1 to N are given according to the depth of the node from its root node).

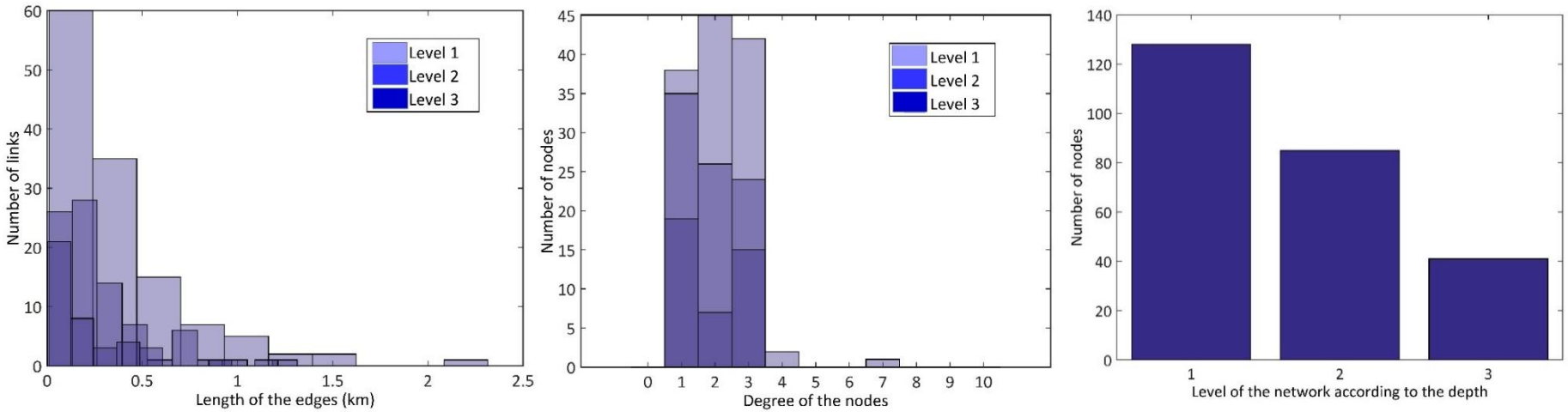
3.7.2 Probability distributions of depth dependent topological properties

An investigation of the depth dependent topological properties was conducted using the same set of real world network data. Since the length and degree related measures play a critical role in describing the topology of a network, the depth dependent degree distributions and the edge length distributions at different depth levels were thoroughly investigated.

Figure 3.12Figure 3.12(a), (b) and (c) show the results of depth dependent analysis of one sub-urban type network (Network 1 in Table 3.2).

From Figure 3.12(a) it can be observed that in Network 1, the edge length distributions of all the three 'levels' approximately follow negative exponential distributions. However, the maximum edge length and the total number of edges in each level has been reduced when going towards level 3 from the level 1 of the network.

Figure 3.12(b) shows the degree distributions of the three levels of the network 1. Figure 3.12(c) shows the distribution of the nodes among the levels of the network 1. It was observed that the number of nodes in each level has been reduced when moving from level 1 to level 3 of the network.



(a) **(b)** **(c)**

Figure 3.12: (a) Depth dependent edge length distribution of Network 1. (b) Depth dependent degree distribution of Network 1. (c) Distribution of nodes among the levels of Network 1.

Table 3.4: Number of nodes in different levels of the networks.

Level	Number of nodes in different levels of sub-urban networks																
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
level 1	128	185	122	150	52	102	93	106	172	39	66	164	94	84	174	180	214
level 2	85	159	109	130	53	80	74	100	107	38	86	117	80	69	104	136	130
level 3	41	54	54	70	31	22	59	40	29	23	44	40	46	20	73	61	40
Total no. of nodes	254	398	285	350	136	204	226	246	308	100	196	321	220	173	351	377	384
Level	**Number of nodes in different levels of urban networks**																
	18	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**28**	**27**	**29**	**30**				
level 1	70	107	175	70	158	171	108	96	106	47	58	74	101				
level 2	96	100	95	99	121	175	112	89	88	44	38	48	121				
level 3	68	30	61	36	49	54	47	42	23	23	19	31	43				
Total no. of nodes	234	237	331	205	328	400	267	227	217	114	115	153	265				

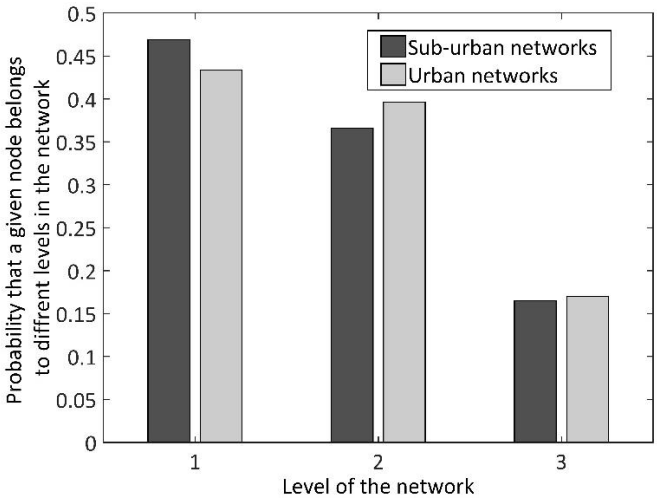


Figure 3.13: Comparison of the distribution of the nodes among the levels of sub-urban and urban networks.

Table 3.4 provides the information regarding the distribution of the number of nodes in different levels of urban and sub-urban networks used in the study. Figure 3.13 compares and summarises the information in Table 3.4. For instance, the average probability that a given node in an urban network belongs to level 1 was obtained by dividing the total number of nodes in level 1 by the total number of nodes in all urban networks. From Figure 3.13 it was identified that in both types of networks the number of nodes in each level has been reduced when going away from the source node. According to Figure 3.13, the fraction of nodes in levels 2 and 3 of urban networks are slightly higher than that of the sub-urban networks. This is due to the higher density of the distribution of consumers in urban areas, compared to that of the sub-urban areas.

Figure 3.14(a) shows a comparison of the edge length distributions of sub-urban and urban networks at their levels 1, 2 and 3. Each curve represents the average variation of all the networks of one type (E.g.: Average variation of 17-sub-urban and average variation of 13-urban networks). When comparing the results, it can be observed that the edge length distributions of all levels in urban networks have a faster decaying negative exponential pattern and a shorter 'maximum edge length' compared to the sub-urban networks. Also, for both types of networks, the maximum edge length that can be observed in level 1 reduced when moving away from the source node towards level 2 and level 3 (Figure 3.14(a)), and the exponential decay of the edge length distribution has also become faster.

Similarly, a comparison of the depth dependent network analysis for the degree distributions is shown in Figure 3.12(b). Compared to the information delivered by the degree distribution curves in Figure 3.8, the depth dependent analysis provides detailed information regarding the network structure. It can be observed that in both types of networks, close to the supply point the network is less branched and when going away from the supply node branching (nodes with degree ≥ 3) and the fraction of leaf nodes (nodes with degree = 1) have increased. Comparison of the degree distribution of the same level in the two types of networks shows that the urban networks tend to have a strong depiction of the above discussed property than that of the sub-urban networks.

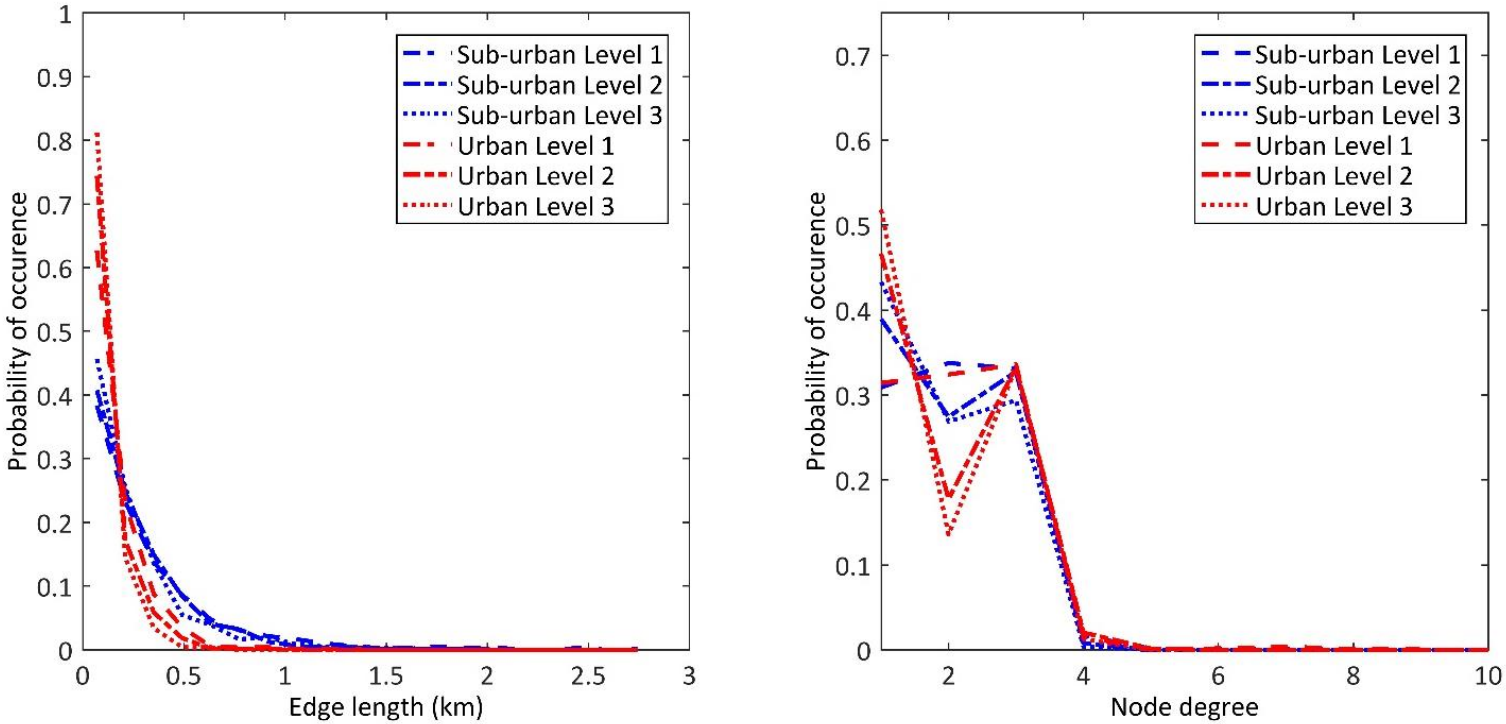


Figure 3.14: (a) Comparison of edge length distributions of sub-urban and urban networks at different levels. (b) Comparison of degree distributions of sub-urban and urban networks at different levels.

Table 3.5: Maximum value for edge length in sub-urban and urban networks at different levels of the network.

Level	Maximum edge length (km)	
	Sub-urban networks	Urban networks
1	2.9	1.4
2	2.0	0.9
3	1.6	0.7

3.8 Summary of the topological investigation

This chapter presents an investigation of the topological properties of real-world electricity distribution networks at the MV level by employing techniques from complex networks analysis and graph theory. A novel approach to obtain depth-dependent topological properties was developed. Topological properties of sub-urban and urban distribution networks are quantified. The sub-urban and urban grouping of the networks is validated using a clustering approach.

Results of the investigation of topological properties in real-world networks showed that,

(i) Node degree and edge length related graph properties are fundamental in characterizing the topological structures of radial type sub-urban and urban electricity distribution networks.

(ii) Depth dependent properties were able to better capture the topological features of electricity networks at different depth levels of the networks. Results from the depth dependent analysis showed that urban and sub-urban types of electricity distribution networks have different graph related properties at different depth levels of the networks.

Chapter 4: Electrical properties of medium voltage electricity distribution networks

4.1 Introduction

Similar to the investigation of topological properties in Chapter 3, it is also important to identify and quantify the electrical properties of different types of real-world distribution networks, in order to produce random-realistic electricity distribution network models.

Some examples for the electrical properties of the distribution networks are Impedances of the distribution lines and electrical equipment, thermal ratings of conductors, protection devices and other electrical equipment, installed capacities of the primary and secondary substations, load profiles of consumers and installed capacities and generation profiles of DGs. In this thesis, electrical properties are referred to as the above electrical parameters and some derivations from one or few of the above electrical parameters (e.g. average installed capacity of a secondary substation, average impedance of a branch) that describe the electrical structure or the electrical arrangement of a distribution network. These electrical parameters are important to study the performance of electricity distribution networks including voltage drops, power losses, network reliability and costs, etc.

The typical assumptions used in these types of studies such as uniform distributions of the consumer load and uniform cross sections and cable types for all the distribution lines are not sufficiently accurate to describe the realistic electrical behaviour of the distribution networks. Distribution of the consumer load is usually different from one network type to the other and

difficult to predict without sufficient information from the real networks. Moreover, the guidelines and recommendations for network planning and design are also different from region to region and cannot be followed exactly in practice due to various reasons such as geographical constrains, policies, regulations and financial factors. Hence, planning and design guidelines of distribution network by its own, is not sufficient to provide realistic representations of the electricity distribution networks.

Therefore, followed by the topological investigation in Chapter 3, an investigation of the electrical properties of real-world, MV electricity distribution networks is carried out in this chapter.

The motivation of this study is to identify and quantify the key electrical properties which will later be useful in network modelling.

The approach of the study is similar to that of the topological investigation in Chapter 3 and the schematic overview of the study is shown in Figure 4.1.

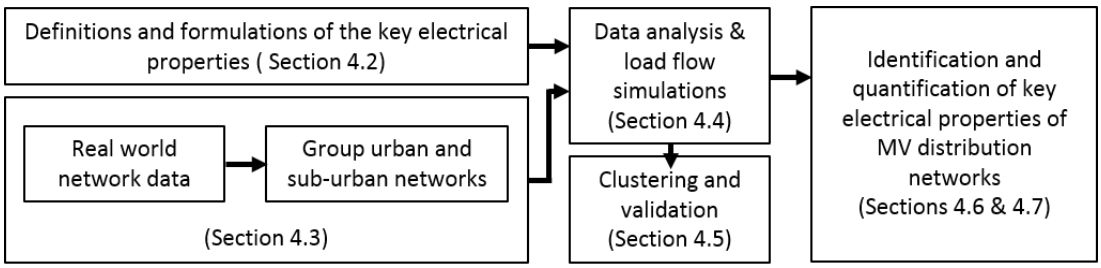


Figure 4.1: Schematic overview of the study.

The same set of Chinese power grid data that was used for the topological investigation, is used for this study. However, only a limited set of electrical data from the real-world networks were available. They are,

(i) Installed capacities of the primary and secondary/distribution substations.

(ii) conductor cross sections of the distribution lines

The following sections of the chapter describe the detailed approach of the investigation of electrical properties of real-world distribution networks using the above available two sets of data.

4.2 Definitions and formulations of the key electrical properties

In order to extract the electrical properties of distribution networks, which are also associated with its topology, each real network is represented as a directed, weighted graph: $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$. $\boldsymbol{V}$ represents the set of nodes and the $\boldsymbol{E}$ represents the set of edges in the graph $\boldsymbol{G}$. The definitions of nodes and edges are similar to those in Chapter 3. Since, all the networks used in this study have radial structures they can be denoted as rooted trees with the primary substation (PS) representing the root *s*, of the corresponding network.

Node identifiers were specified according to the Breadth-first search (BFS) approach [70]. In this approach, node numbering starts at the tree root ($node\_id$ equals to 1), and number the nodes in the next same depth level first (E.g.: $node\_id$ = 2, 3, 4...), before moving to the subsequent depth level. Lengths (in km) of the distribution line segments are used as the weights of the edges in the graph.

Definitions and formulations of the key electrical properties used in this study are described in the following section with reference to the 10kV Chinese electricity distribution networks selected for the investigation. Table 4.1 provides a summary of the electrical properties investigated in this study.

Table 4.1: Summary of the electrical properties.

Category	Electrical properties
Substation capacity related properties	Installed capacity of the primary substation (MVA)
	Average installed capacity of a secondary substation (kVA)
	Variation of the installed capacities of the secondary substations with the distance to the source node along the feeders
	Load density(MVA/km^2)
Conductor cross section related properties	Average electrical distance(Ohms) between two nodes of the network (i.e. average resistance and reactance between two nodes in the network)
	Thermal ratings of the conductors (kVA or amperes)
Electrical performance of the network related properties	Total power loss to the total supplied power ratio (%) of the network
	Minimum recorded voltage of the network
	System load balancing index

4.2.1 Substation capacity related properties

- **Primary substations (PS)**

Usually, a 10kV distribution network is supplied by a PS which transforms voltage from 35kV to 10kV. PSs are often sited close to large industrial customers or at the load centre. In the graph representations of electricity networks, PS becomes the root node.

The installed capacities of these primary substations (PS_c) depends on the type of the area it is located. In rural areas the primary substation capacity can be quite small with a single 35kV/10kV transformer and no more than 2 or three outgoing feeders. A feeder in an electricity distribution networks is an exit route that carries electric power from the primary substation to the distribution substations. Primary substations in city centre areas often have several large transformers and more than 10 outgoing feeders. The information of PSs are specified in the following format (4.1).

$$PS = [node\_id, \quad PS_c\ (MVA), \quad num\_of\_feeders]_{1\times3} \tag{4.1}$$

where, $node\_id$ is the identifier of the node where the PS is located, PS_c is the installed capacity of the PS and $num\_of\_feeders$ is the number of outgoing feeders from the PS.

- **Secondary substations (SS)**

The secondary substations (SS), where power is transformed from 10kV down to the 0.4kV LV system come in different types. In highly populated areas secondary substations are mostly ground mounted and have higher installed capacities than in less populated areas. In less populated areas pole mounted substations are widely used.

In graph models of the distribution networks, each node representing a secondary substation is assigned with an installed capacity, SS_C. The distance (in km) from root *s*, to the secondary substations along the feeders ($dist_{s,SS}$), is obtained using the Djakarta's shortest path algorithm [65]. The list of SS is represented in the matrix format as shown in Equation (4.2). N_{SS} is the total number of secondary substations in the network.

$$SS \quad = [node\_id, \quad dist_{s,SS}\ (km), \quad SS_C(kVA)]_{N_{SS}\times3} \tag{4.2}$$

- **Average installed capacity of secondary substations**

The average installed capacity of secondary substations ($SS_{C,avg}$) in a network can be obtained from the above data using Equation (4.3). A High value of $SS_{C,avg}$ implies a high electrical load density and a lower value of $SS_{C,avg}$implies a low load density in a selected area of the networks.

$$SS_{C,avg} = \frac{\sum_{i=1}^{N_{SS}} SS_{C,i}}{N_{SS}} \tag{4.3}$$

- **Distribution of the capacities of secondary substations along the feeder lengths**

The distribution of consumer loads is a primary factor that determines the distribution of SS capacities. The distribution of consumer loads can be different from one network type to the other. For example, in rural type distribution networks, usually the PSs are located closer to the high capacity loads while some distribution feeders stretch further to provide power to the far away low capacity loads.

In order to capture the variation of SS capacity with its distance to the root node, the following approach is used. The SS capacities are plotted against the distance to the root node of each SS. The distances are normalised over entire network (l_{norm}) by dividing the distances to each SS from the root node, by the maximum feeder length l_{max}, of the corresponding network as shown in Equation (4.4). l_{max} of a network is the distance from root node to the furthest away SS along a feeder.

$$l_{norm} = \frac{dist_{s,SS}}{l_{max}} \tag{4.4}$$

This normalisation allows the comparison of the distribution of SS capacities along the feeder length between different networks.

- **Load density (MVA/km^2)**

The load density, LD (MVA/km^2) of a given area of the network is obtained by dividing the total installed capacity of SSs in the network by the total supply area ($Total\ area$) of the network and then multiplying this value by a demand factor k_d as shown in Equation (4.5). A demand factor k_d , for this study is defined as a fraction of the installed capacity of a SS in the distribution network, that is actually being consumed as the consumer load at a given instant of time. In the

real-world networks SSs do not operate in their maximum capacity under normal operating conditions. Therefore, it was assumed that on average SSs are operating at 60% of their capacities and hence, for the calculations, k_d is assumed to be equal to 0.6.

$$LD = \frac{\sum_{i=1}^{N_{SS}} SS_{C,i}}{Total\ area} \times k_d \tag{4.5}$$

4.2.2 Conductor cross section related properties

Usually, in radial feeders, the distribution line segments closer to the root node carry the maximum power while the line segments further away from the source are lightly loaded (Note: this explanation is valid when there is no DG is connected to the distribution network. With DGs connecting at various places the line loadings in the distribution networks can be different from the above explanation). These characteristics of radial feeders enable the choice of multiple conductor cross sections for a single feeder. However, the use of a large number of conductors of different cross sections will result in an increased cost of the inventory [71]. To avoid high costs, specific recommendations are set by the network planners.

The distribution line segments in a network can be divided into a trunk and lateral branches (tie lines are ignored). The trunk refers to the backbone of a network and is the main route of power transfer from the primary substation down to the load centres. Usually, the trunk lines have larger cross sections and also a headroom is left for future expansions. Lateral branches are the distribution line segments that connect between trunk lines and the MV consumers and SSs.

The choice between overhead lines (OH) or underground (UG) cables for a distribution line depends on the type of area. UG cables are widely used in urban areas while OH lines are more common in rural or less populated areas. In addition, the choice of UG cables and OH lines are

also heavily influenced by the cost of labour and material, and geographical constrains such are rivers, mountains, reserved areas, etc. [66].

- **Impedance of a distribution line**

Depending on the conductor cross section (mm^2), the per km impedance (z) of a distribution line can be obtained using manufacturer manuals. Parameter r is the per km resistance and x is the per km reactance in Equation (4.6). Usually the shunt capacitance of the distribution lines is ignored [66].

$$z = r + jx \tag{4.6}$$

- **Electrical distance matrix**

If the impedances of all the distribution lines are known, an electrical distance matrix (ED), for the network can be obtained as shown in Equation (4.7),

$$ED_{ij} = \begin{cases} d(i,j) \times z(i,j) , & if (i,j) \in E \; and \; i < j \\ 0, & otherwise \end{cases} \tag{4.7}$$

where, $d(i,j)$ is the length of the distribution line connecting nodes i and j and $z(i,j)$ is the impedance of the distribution line connecting nodes i and j.

- **Average electrical distance between two nodes in a network**

Average electrical distance ed_{avg}, between the nodes in the network is then obtained according to Equation (4.8). N is the total number of nodes in the network and M is the total number of branches in the network.

$$ed_{avg} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} ED(i,j)}{M} \quad (4.8)$$

- **Thermal rating of the conductors**

Thermal rating/limit of the conductors is one major factor that limits the power flow through a distribution line. These thermal limits are intended to limit the temperature reached by the energized conductors and the resulting sag and loss of tensile strength. Thermal ratings (in terms of kVA or amperes) of the corresponding cross section and material type of the conductors can be found using cable manufacturers data manuals.

4.2.3 Electrical performance related properties

In practice, electricity distribution networks are designed to meet many technical and economic standards to achieve the best performance in both normal and abnormal operating conditions. Power losses, voltage drops and feeder utilizations are some of the most important indicators to describe the electrical performance of a distribution network.

- **Power losses**

Given the active (P) and reactive (Q) load connected at the end of a distribution line segment e_{ij} with a conductor impedance (z), the active power loss ($PLoss$), reactive power loss ($QLoss$) and apparent power loss ($SLoss$) is obtained by the following Equations (4.9), (4.10) and (4.11).

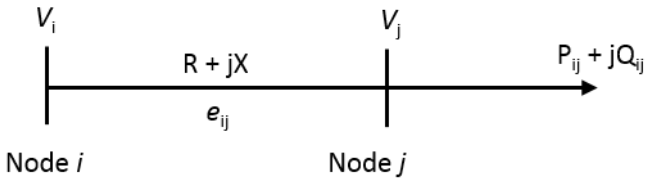


Figure 4.2: Representation of a distribution line segment.

$$PLoss_{e_{ij}} = \frac{P_{ij}^2 + Q_{ij}^2}{V_j^2} \times R = \frac{S_{ij}^2}{V_j^2} \times R \tag{4.9}$$

$$QLoss_{e_{ij}} = \frac{P_{ij}^2 + Q_{ij}^2}{V_j^2} \times X = \frac{S_{ij}^2}{V_j^2} \times X \tag{4.10}$$

$$SLoss_{e_{ij}} = \sqrt{(PLoss_{eij}^2 + QLoss_{eij}^2)} \tag{4.11}$$

The total power loss $SLoss_{Total}$ is the sum of individual power losses in each line segment in the network.

The ratio of total power loss to the total supplied power ($Power\ loss\ ratio$) is used as another performance evaluation parameter to compare different networks. $Ssupplied_{Total}$ is the total power supplied to the network.

$$Power\ loss\ ratio = \frac{SLoss_{Total}}{Ssupplied_{Total}} \times 100\% \tag{4.12}$$

- **Voltage drops**

In a distribution network, a knowledge of the voltage at different parts of the network can indicate the strong and weak parts of a network. Often a specific voltage drop (E.g. 6%) is permitted between the supply node and any other node in the network. The voltage drop V_d in a distribution line segment can be calculated using the following formula:

$$V_d = \frac{P}{V}(R + X\ tan\varphi) \tag{4.13}$$

where, V is the rated line-line voltage, P is the total three phase active power and φ is the angle between voltage and current. In the above Figure 4.2, the load is concentrated at the receiving end. The total voltage drop V_{d_Total} is the sum of individual voltage drops due to each load point [66].

- **Load balancing index**

Branch load balancing index LB_{branch} and overall system load balancing index LB_{sys} are used to determine the loading condition of the distribution network. Branch load balancing index is defined as a measure of how much a branch can be loaded without exceeding the rated capacity of that branch. This can be represented mathematically as,

$$LB_{branch} = \frac{S_{branch}}{S_{branch}^{max}} \tag{4.14}$$

where, S_{branch} is the complex power flowing through the branch and S_{branch}^{max} is the maximum rating or capacity of branch. The system load balancing index LB_{sys} of the entire distribution network is represented as,

$$LB_{sys} = \frac{1}{M}\sum_{branch=1}^{M}\frac{S_{branch}}{S_{branch}^{max}} \tag{4.15}$$

where, M is the total number of branches in the network [72], [73].

4.3 Power grid data for the investigation of electrical properties

Sub-urban and urban network samples with radial structures were selected from the data of 10kV Chinese electricity distribution networks. Similar to the topological investigation (Section 3.3), network classification as sub-urban and urban is done according to the population density. Table 4.2 summarises the basic information of the 10kV network samples selected for this study.

Table 4.2: Basic information of the selected samples of 10kV sub-urban and urban networks.

Network type	Network ID	Area (km^2)	Population density ($/km^2$)	Total number of nodes	Total Number of branches	Number of outgoing feeders from the primary substation	Installed capacity of the primary substation (MVA)	Number of secondary substations in the network	Total installed capacity of secondary substations (kVA)	Maximum feeder length (km)
Sub-urban	1	66.2	482	254	253	7	8+5 = 13	92	16100	10.7
	2	142.0	328	399	398	5	2×10 = 20	129	20652	14.8
	3	64.5	405	285	284	4	2×10 = 20	109	20083	10.8
	4	80.9	449	173	172	4	2×10 = 20	65	13470	12.3
	5	108.5	256	204	203	5	2×8 = 16	81	17303	11.8
	6	78.5	369	196	195	4	2×10 = 20	76	15135	11.5
	7	84.1	473	321	320	6	2×10 = 20	117	24440	14.4
	8	88.0	508	351	350	7	2×10 = 20	109	28080	15.1
	9	32.2	422	169	168	6	5+6.3= 11.3	59	12580	6.8
	10	48.8	640	186	185	5	2×10 = 20	73	17650	7.1
Urban	11	9.8	1939	234	233	9	2×20 = 40	65	23850	3.8
	12	9.2	1935	237	236	7	2×50 = 100	98	50030	3.7
	13	10.0	1930	331	330	7	2×16 = 32	90	38200	3.5
	14	8.0	1750	205	204	7	50+63= 113	83	43900	5.2
	15	14.0	3200	328	327	8	2×50 = 100	125	72370	5.5
	16	16.0	3200	400	399	7	2×20 = 40	115	66380	4.8
	17	7.0	3600	227	226	6	31.5+20 =51.5	99	42250	4.0
	18	7.5	3600	217	216	6	2×50 = 100	96	42865	5.6
	19	8.0	3600	265	264	10	2×20 = 40	89	65080	5.1
	20	7.5	3600	153	152	7	2×16 = 32	50	27470	3.4

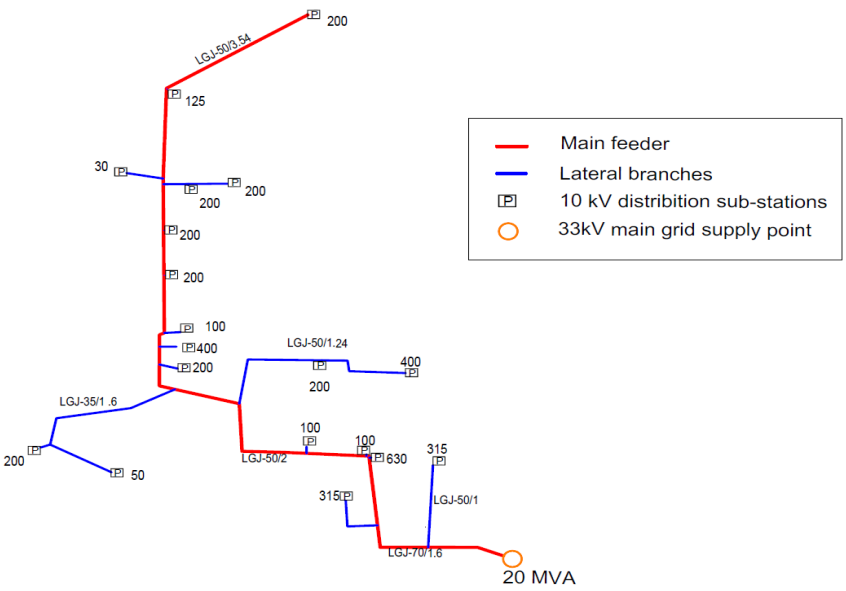


Figure 4.3: A distribution feeder of a 10kV electricity distribution network.

Figure 4.3 shows an example for a real distribution feeder of a 10kV electricity distribution network. The trunk line and the lateral branches are marked in red (thick) and blue (thin) lines. Installed capacities of the 10kV/0.4kV secondary substations are marked next to each SS. The capacity of the 35kV/10kV primary substation is 20MVA. Conductor type used for the distribution lines are marked along the lines. Consider the example of conductor type LGJ-70/1.6 in Figure 4.4. 'LGJ' is the symbol of steel cored aluminium strand, in which 'L' the abbreviation of aluminium wire is, 'G' is the abbreviation of steel core, and 'J' is the abbreviation of stranded wire. The number '70' refers to the cross section (mm^2) of the conductor [74]. In the network drawings the number after the cross section (e.g. 1.6) refers to the length of the distribution line segment with that conductor cross section.

Table 4.3 summarises the installed capacities of the SSs and PSs of the selected set of sub-urban and urban network samples. It was noticed that, at some primary and secondary substations, two or more transformers were installed to obtain a higher power supply capacity or to act as a backup unit.

Table 4.3: The selection of capacities for primary and secondary substations transformers in 10kV networks.

Substation type	Unit	Installed capacities
Primary substation PS (33kV/10kV)	MVA	5, 6.3, 8, 10, 16, 20, 31.5, 50, 63
Secondary substation SS (10kV/0.4kV)	kVA	20, 30, 50, 80, 100, 125, 160, 200, 250, 315, 400, 500, 625, 800,1000

Table 4.4 summarises the properties of widely used conductor types in the selected set of 10kV distribution networks.

Table 4.4: Properties of the conductor types used for 10kV distribution lines [6].

Cable name	Cross section mm^2	Resistance per km (Ohms)	Reactance per km (Ohms)	Current rating at 90 degrees (A)
LGJ-35	35	0.82	0.38	180
LGJ-50	50	0.59	0.368	227
LGJ-70	70	0.42	0.358	287
LGJ-95	95	0.29	0.342	338
LGJ-120	120	0.23	0.335	390
LGJ-185	185	0.16	0.365	518
LGJ-240	240	0.12	0.358	610
LGJ-300	300	0.09	0.365	707

4.4 Quantification of the electrical properties

In order to compute electrical performance related network parameters load flow simulations were conducted. However, the consumer load profiles or load data were not available in the original data. To overcome this limitation a few assumptions were made. A steady state power flow simulation for each network was conducted with the following assumptions.

(i) A fraction (60%) of the installed capacities of secondary substations are used as the actual load connected to each load node.

(ii) Power factor at each load node is considered as 0.9.

(iii) Per unit voltage at the supply point is equal to 1.05 p.u.

Table 4.5 presents the results of the evaluation of electrical properties of the above sub-urban and urban networks at 10kV level.

In order to compare the results in Table 4.6, probability distributions of the electrical properties of both sub-urban and urban networks were obtained. Figure 4.4 shows the comparative probability distribution plots for the two types of networks, arranged back to back on the *x*-axis (probability of occurrence). For one electrical property, the same bin size and the same number of bins were used to generate the probability distributions of both types of networks.

Table 4.5: Electrical properties of the 10kV sub-urban and urban networks.

Network ID	Load density (MVA/km^2)	Average installed capacity of a secondary substation (kVA)	Average Electrical distance		Total power loss/total supplied power ratio (%)	Minimum recorded voltage (pu)	System load balancing index
			R (Ohms)	X (Ohms)			
1	0.146	175	0.2566	0.1129	4.84	0.949	0.1257
2	0.145	160	0.1167	0.0798	5.85	0.912	0.1568
3	0.187	184	0.0874	0.0798	6.13	0.954	0.11
4	0.166	207	0.2571	0.1179	3.83	0.983	0.1134
5	0.159	213	0.3364	0.1467	5.43	0.948	0.1651
6	0.116	199	0.2515	0.1176	5.93	0.927	0.136
7	0.174	209	0.1504	0.0986	4.42	0.964	0.144
8	0.191	258	0.1071	0.0965	7.99	0.896	0.1294
9	0.234	213	0.2757	0.122	3.8	0.962	0.1448
10	0.217	242	0.2409	0.1138	2.98	1.004	0.142
11	1.46	367	0.0886	0.0449	2.17	1.027	0.1496
12	3.263	511	0.0511	0.0446	4.14	0.970	0.1834
13	2.292	424	0.0456	0.037	1.4	1.033	0.0938
14	3.293	529	0.0715	0.0596	2.42	1.017	0.1716
15	3.102	579	0.0405	0.0391	5.11	0.979	0.3047
16	2.489	577	0.0519	0.0407	5.03	0.966	0.3479
17	3.621	427	0.0613	0.0412	2.09	1.012	0.2169
18	3.429	447	0.0584	0.0444	6.22	0.949	0.2844
19	4.881	731	0.0479	0.0455	4.21	0.994	0.2771
20	2.198	549	0.0742	0.0541	3.05	1.015	0.2526

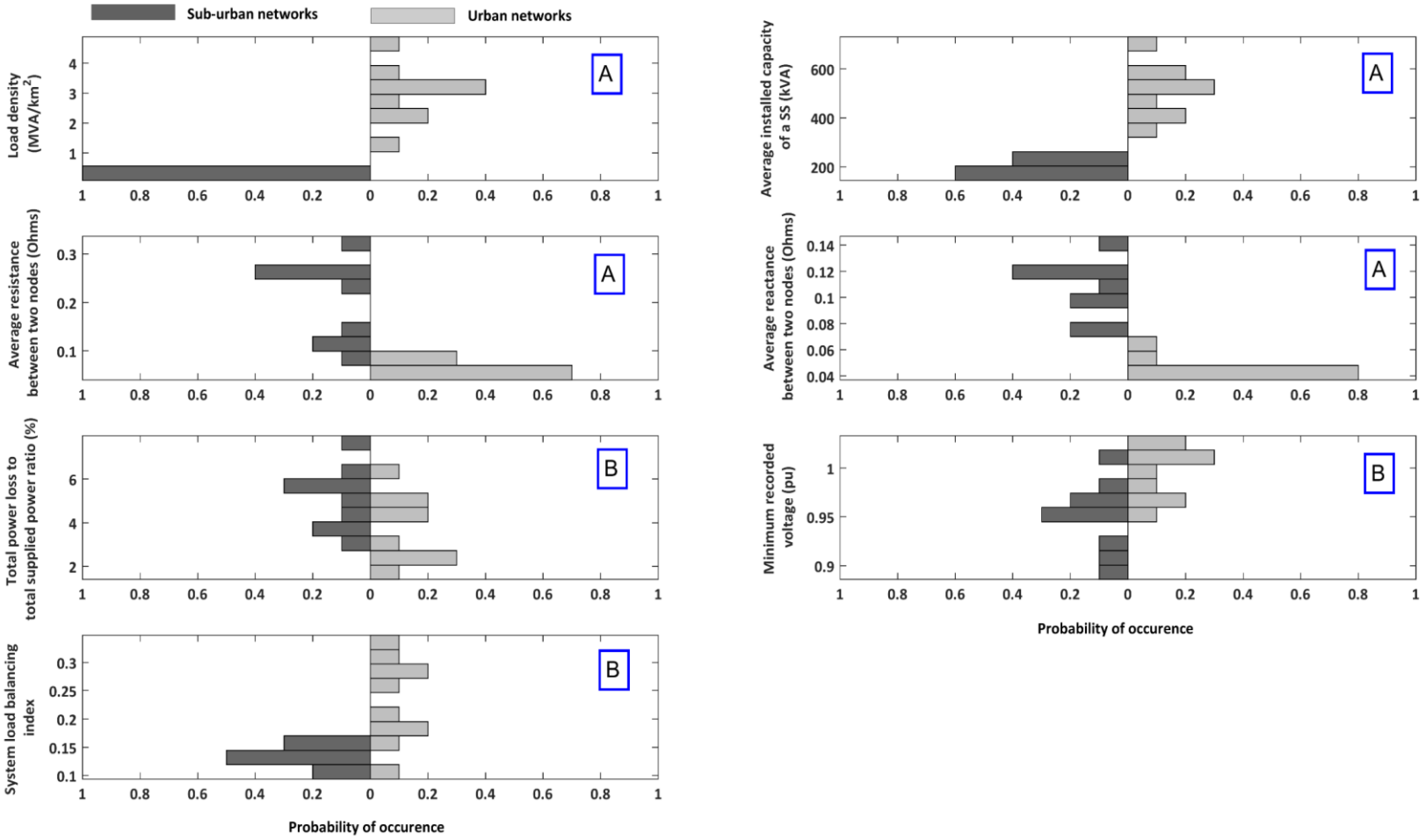


Figure 4.4: Comparison of the electrical properties of urban and sub-urban networks.

From Figure 4.4 it can be identified that some of the electrical properties are able to clearly characterise the two types of networks. These properties include load density, average installed capacities of secondary substations, average resistance and reactance between the nodes of the network (sub-graphs with letter A).

The ranges of variation of these properties in sub-urban and urban networks are summarised in Table 4.6. In sub urban networks, load densities and the average installed capacities of SSs are considerably lower than that of the urban networks. In the real situation these observations from the results are reasonable, as the urban areas are highly populated and a higher number of commercial and industrial consumers are located in the urban areas than in the sub-urban areas. However, this study has managed to quantify these characteristics of sub-urban and urban networks in terms of load densities and average capacities of SSs and clear differences between the ranges of variations of the values have been identified.

Table 4.6: The ranges of variations of the electrical properties of sub-urban and urban distribution networks.

Electrical property	**The range of variation**	
	Sub-urban networks	Urban networks
Load density (MVA/km^2)	0.116 - 0.234	1.460 - 4.881
Average installed capacity of secondary substations (kVA)	160 - 258	367 - 731
Average resistance between two nodes of the network (Ohms)	0.087- 0.336	0.041 - 0.089
Average reactance between two nodes of the network (Ohms)	0.080 - 0.147	0.037 - 0.060

According to the summary in Table 4.6, average resistance and reactance of the edges connecting two nodes of the urban networks are lower than that of the sub-urban networks. This observation from the results is explained by two factors of the real-world networks; (i) as studied in Chapter 3 about the average edge length between two nodes in the sub-urban and

urban networks, urban networks have shorter average edge lengths between two nodes of the network than in the sub urban networks (ii) conductors used in the urban networks for distribution lines have larger cross sections and hence lower resistances and reactances.

The probability distributions of the properties labelled with letter 'B' in Figure 4.4 also have some noticeable differences (Probability distribution plots of the two network types are biased into different directions) between the two types of networks. It can be observed that the Load balancing index and minimum recorded voltages of urban networks tend to be higher than the sub-urban network while the total power loss to total supplied power ratio in sub-urban networks tend to be higher than that of the urban networks. However, there are still some overlapping of the values. Therefore, these properties are not suitable to clearly characterise the two types of networks.

4.5 Validation of the networks classification using electrical properties

A clustering exercise following the same clustering approach described in the topological investigation (Section 3.5) was carried out with the results in Table 4.5 for electrical properties of the networks.

First, all the 7 electrical properties mentioned in Table 4.5 were chosen as the cluster variables. The *k*-means algorithm was used to group the 20 networks into two clusters. The grouping done by using all the 7 variables, exactly followed the sub-urban and urban classification done by the population density parameter of the networks. Hence this step was used to validate the urban, sub-urban classification of the networks by using the electrical properties.

The results shown in Figure 4.5 are graphical representations of the case where, different subsets with 3 parameters were chosen as the cluster variables. In the first two cases (Figures 4.5(a) and 4.5(b)), the selected sets of cluster variables were able to group the network sample into two clusters accurately, as defined by the population density of the networks. However, the third set of cluster variables shown in Figure 4.5(c) did not cluster the network sample into the right groups. Similar to that of the topological investigation, this observation explains the importance the careful selection of network properties when characterizing different network types.

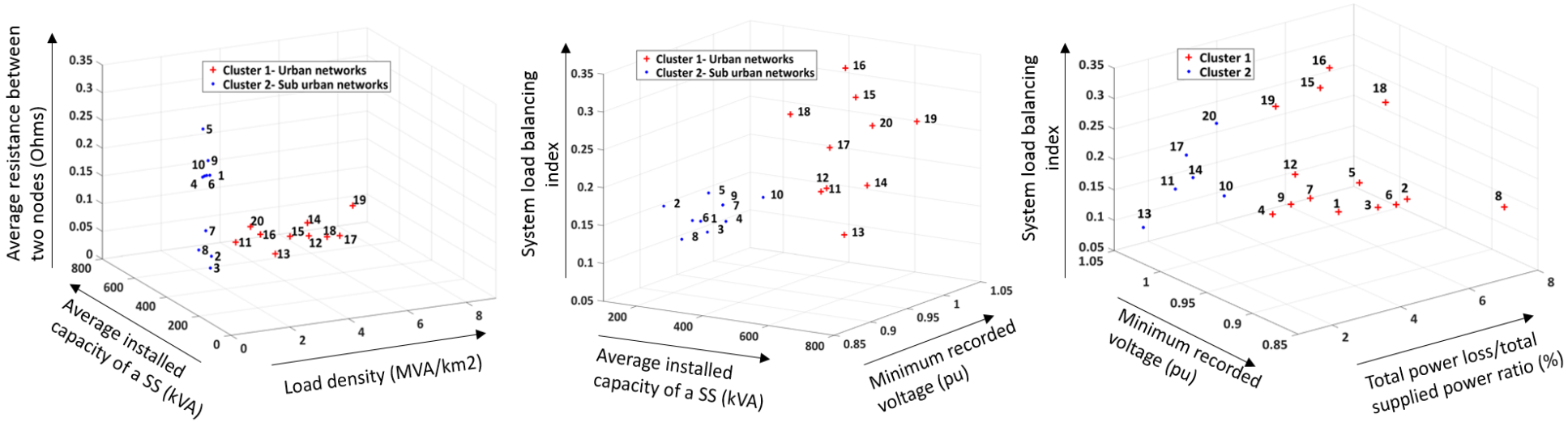


Figure 4.5: Cluster assignments.

4.6 Identification of the key electrical properties for characterising electricity distribution networks

From the results, it is evident that the electrical properties derived from the data of installed capacities of the secondary substations and the conductor cross sections used for distribution lines are able to characterise the different types of networks.

The distribution of the installed capacities of the secondary substations is an indirect representation of how the consumers are distributed in a network, which is not possible to be derived from the planning and design guidelines. Therefore, realistic allocation of the installed capacities of the substations on a network is important in achieving realistic representations for different types of electricity distribution networks.

The allocation of the conductor cross sections of a sub-urban type real world distribution network is shown in Figure 4.6. Similar to the network in Figure 4.6, in all the real-world network samples selected for this study, the trunk lines are assigned with larger cross sections while the laterals are assigned with slightly smaller conductor cross sections. Also, a maximum of 3 - 4 different types of conductor cross sections were observed in each network used for this study, which also is consistent with the normal practices of network planning and design [75].

In the network modelling point of view, further investigating the distribution of the installed capacities of the secondary substations has clear benefits. If the realistic distributions of secondary substations can be obtained the other electrical parameters to the representative topology models can be assigned by following the traditional approach for planning and design of distribution network.

The traditional approach for planning and design of distribution networks involves forecasting of electricity demand, locating and sizing substations, designing the layout of the power distribution network and finally the selection of electrical equipment (E.g.: Selection of conductor cross sections for OH lines, UG cables) [66].

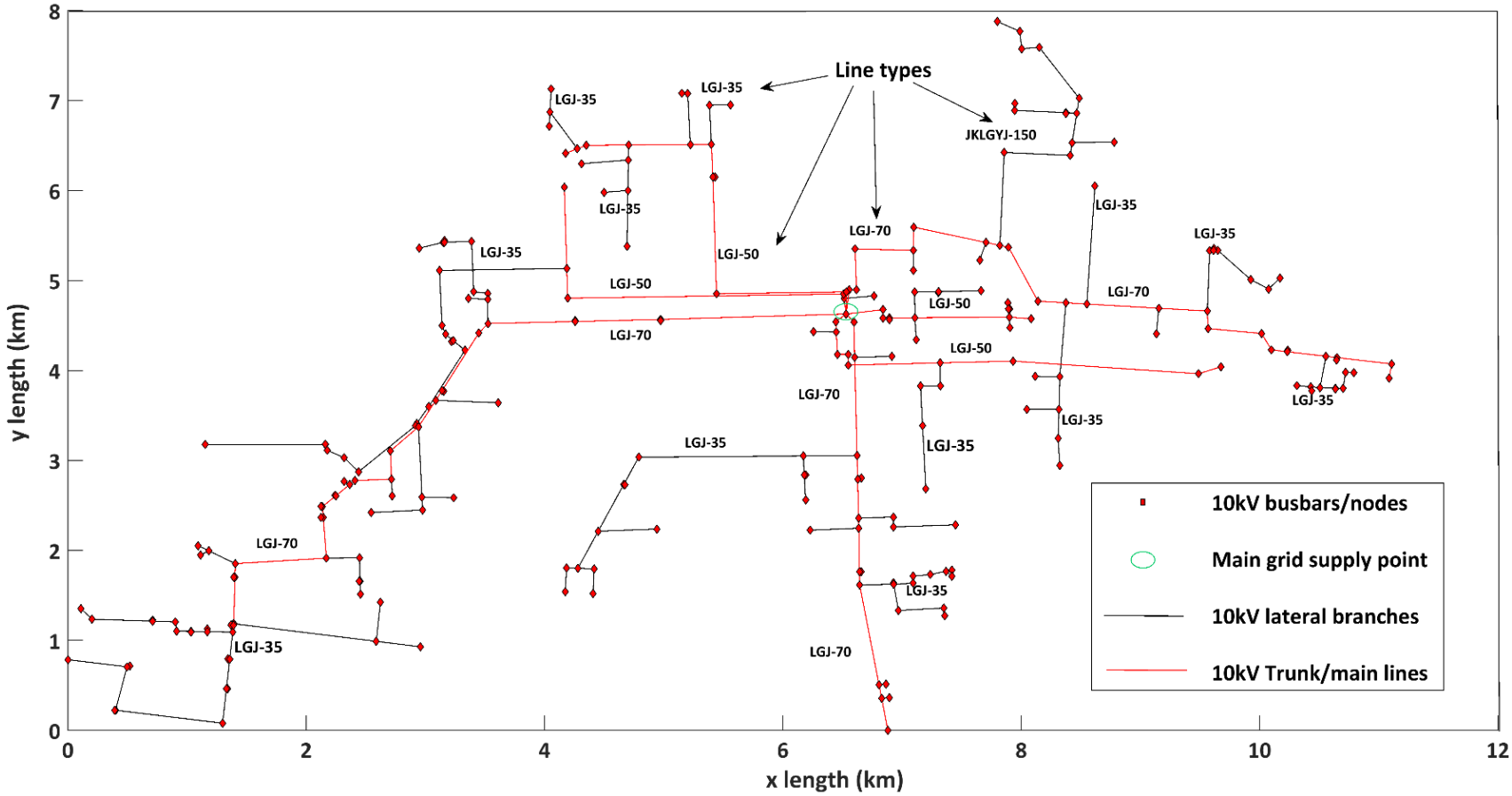


Figure 4.6: Network layout and the selection of conductor cross sections of the sub-urban type network no.1.

4.7 Network depth dependent electrical properties

In this section the depth dependent electrical properties are investigated. The depth in this study is the normalised distance, l_{norm} from the root node to the other nodes in the network (Equations (4.4)). Whereas in the topological investigation in Chapter 3, depth dependent properties were calculated considering the depth as the number of steps from the source node to the other nodes in the network (Equations (3.18) and (3.19)). The aim is to capture the realistic distribution of secondary substation capacities of sub-urban and urban networks.

Figures 4.7(a) and 4.7(b) show the distribution of the installed capacities of SSs in sub-urban and urban networks with the normalised distance from root node to each SS. The information in Figure 4.7(a) corresponds to all the sub-urban networks used in the study and similarly Figure 4.7(b) corresponds to all the urban networks.

The range of variation in SS capacities in urban networks is much larger compared to that of the sub-urban networks. In the sub-urban networks, the number of SSs with higher installed capacities are less at the far ends of the network from the root node. Whereas in urban networks, high capacity SSs are spread almost evenly throughout the lengths of the networks

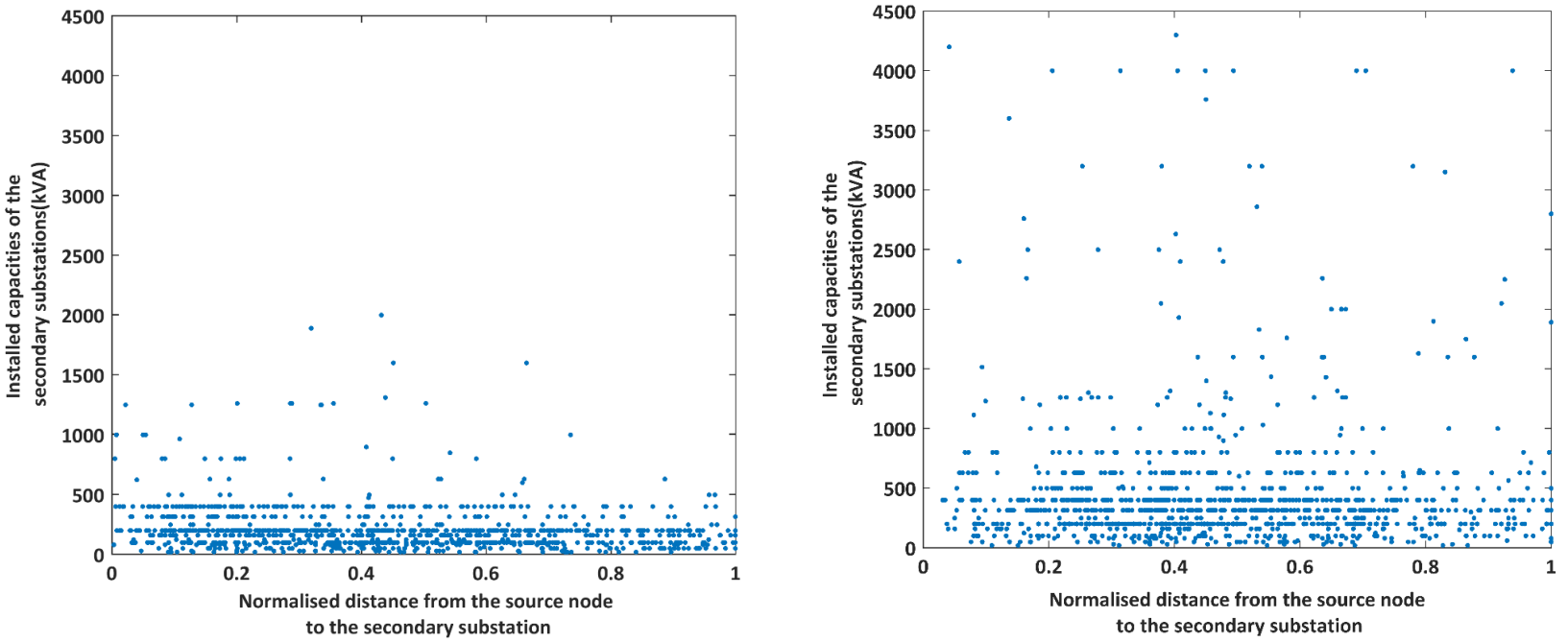


Figure 4.7: Distribution of the installed capacities of secondary substations with the normalised distances from the source nodes in, (a) Sub-urban networks (b) Urban networks.

4.7.1 Definition of 'depth levels' for the analysis of electrical properties

In order to better understand the distribution of the secondary substation capacities, the networks are divided into levels according to the normalised lengths, l_{norm} from the source node as shown in Figure 4.8. For example, secondary substations within the normalised distance 0 and 0.1 belong to the $level_1$ of the network. The number of levels (m) can be varied. 10 levels are chosen for this study. A sensitivity study with different number of levels can be conducted to identify the best number of levels to capture the depth dependent properties of different types of networks. However, this part of the study is not covered in this thesis.

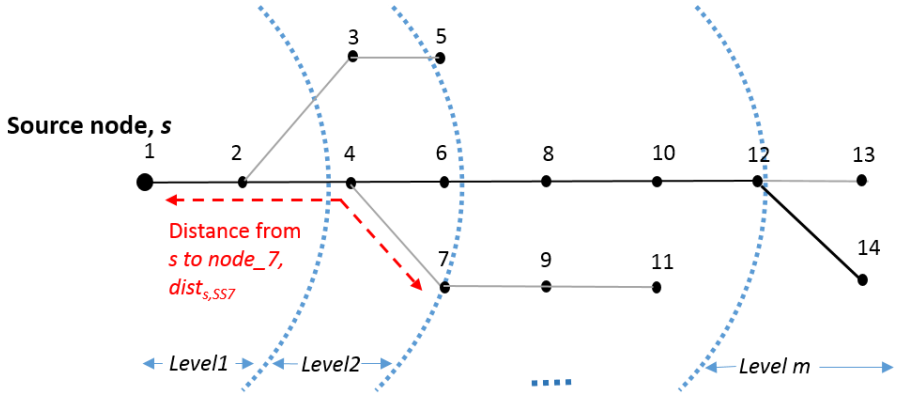


Figure 4.8: Representation of levels of a network for the analysis of electrical properties.

$$level = \begin{cases} level_1; & 0 < l_{norm} \leq 0.1 \\ level_2; & 0.1 < l_{norm} \leq 0.2 \\ & \vdots \\ level_{10}; & 0.9 < l_{norm} \leq 1 \end{cases} \quad (4.16)$$

4.7.2 Probability distributions of the depth dependant electrical properties

Applying the Kernel distribution fitting methodology described in [76], the PDFs of the installed capacities of SSs at different levels of the two types of networks are calculated (Appendix A). Unlike a histogram, which discretizes the data values into separate bins, the Kernel distribution

sums the weight functions for each data value to produce a smooth, continuous probability curve.

The Kernel density probability distributions for the installed capacities of SSs in each level can be formulated using Equations (4.17) and (4.18).

$$f_h(y) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{y - y_i}{h}) \tag{4.17}$$

$$K(y) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}} \tag{4.18}$$

With relevant to this study, variable y in Equations (4.17) and (4.18) refers to the installed capacities of the SSs in the given level of the networks and n refers to the total number of SSs in that level. The Kernel distribution is defined by a kernel smoothing function $K(y)$ and a bandwidth value *h* that controls the smoothness of the resulting density curve. In this model, the normal Kernel smoothing function is used, described in Equation (4.18). The bandwidth value h is considered to be the optimal for estimating densities for the normal distribution.

Basic statistics such as mean and standard deviation of the probability distributions can be used to describe the shape of the PDFs derived from the above Kernel density estimation [77]. In probability and statistics, the mean value, $\overline{y}$ (i.e. is same as the expected value of y, $E(y)$) is referred to the central tendency of a random variable, y characterised by a probability distribution, $f_h(y)$ as shown in Equation (4.19). Standard deviation, std is a measure that is used to quantify the amount of variation of a set of data values from the mean, as shown in Equation (4.20). Here the operator E denotes taking the expected value of a random variable.

$$\overline{y} = E(y) = \int_{-\infty}^{\infty} y\, f_h(y) \tag{4.19}$$

$$std = \sqrt{E((y - \overline{y})^2)} \tag{4.20}$$

Figures 4.9 and 4.10 show the resulted PDFs of the Kernel density estimation for SS capacities of sub-urban and urban networks respectively. SS capacities are positive values. Therefore, only the positive values of the x-axis are shown in the figure. 'Standard deviation', is shown only on one side of the 'mean' due to this reason.

From Figure 4.9 it can be observed that, the mean values of the PDFs of sub-urban networks are gradually decreasing when moving down from level 1 to level 10. This means, in sub-urban networks the higher capacity SSs are installed closer to the PS and the capacities of the SSs which are further away from the PS tend to have smaller values. The standard deviation of the PDFs has the highest deviation from the mean at the middle levels of the network. This explains that in the middle depth levels of the sub-urban networks the SS capacities vary within a large range of values from very high capacities to low capacities. However, the standard deviation of the PDFs at the far ends of the networks is much smaller, proving that the installed capacities of SSs at the far ends of the networks vary between a small ranges of values.

In the urban networks (Figure 4.10) the variation of the mean and standard deviations when going down from level 1 to level 10, does not have a clear pattern. However mean values do not vary largely from one level to the other (variation is around 500kVA). These observations of the changes in mean and standard deviation explains that, in general, the capacities of SSs in the urban networks are distributed almost evenly throughout all the depth levels of the networks.

However, standard deviations and means of the PDFs in urban networks are much higher than that of the sub-urban networks for all the depth levels in the networks. Results show a clear difference of the SS capacity distribution among sub-urban and urban networks.

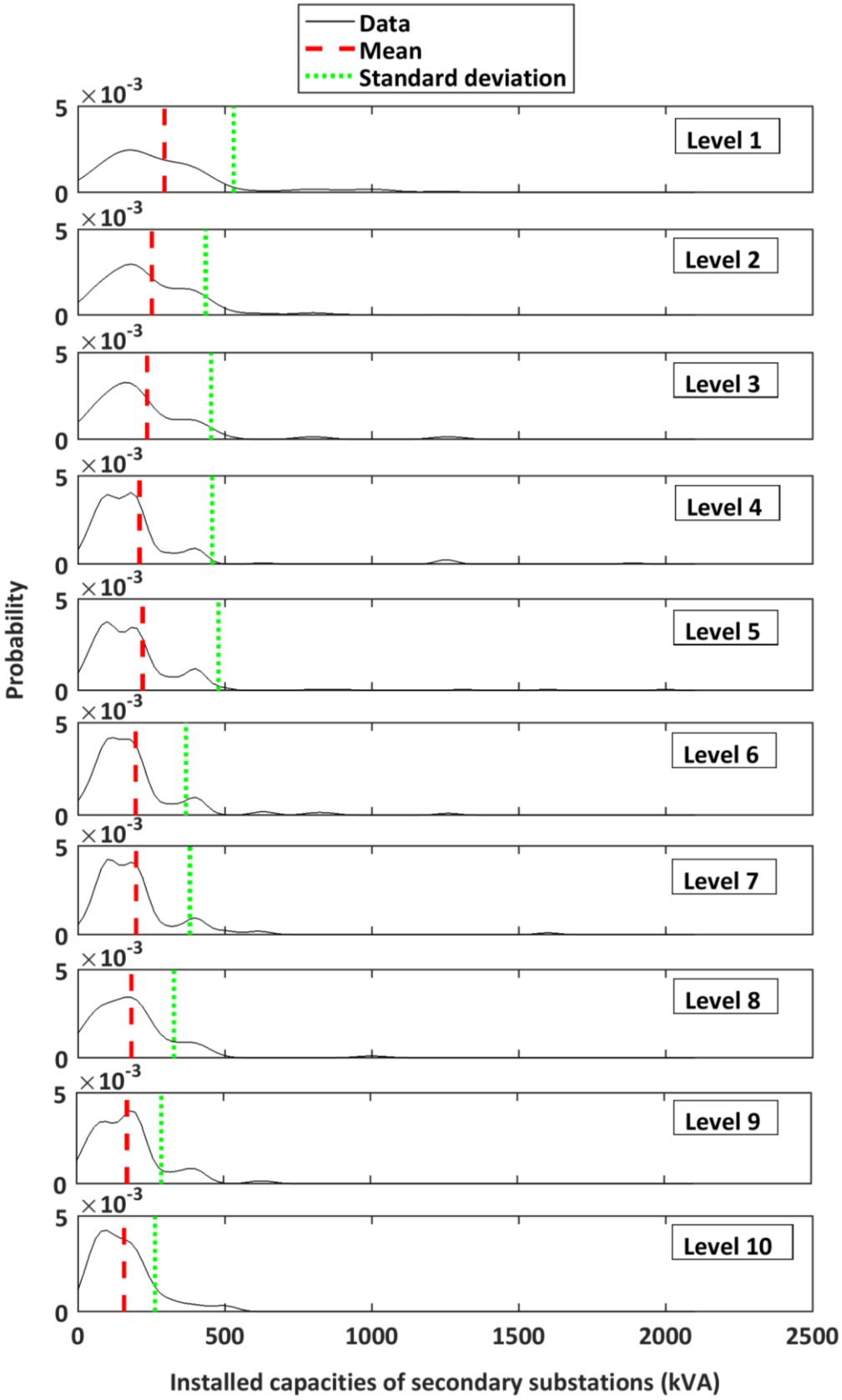


Figure 4.9: Level dependent distribution of the installed capacities of secondary substations in sub-urban networks.

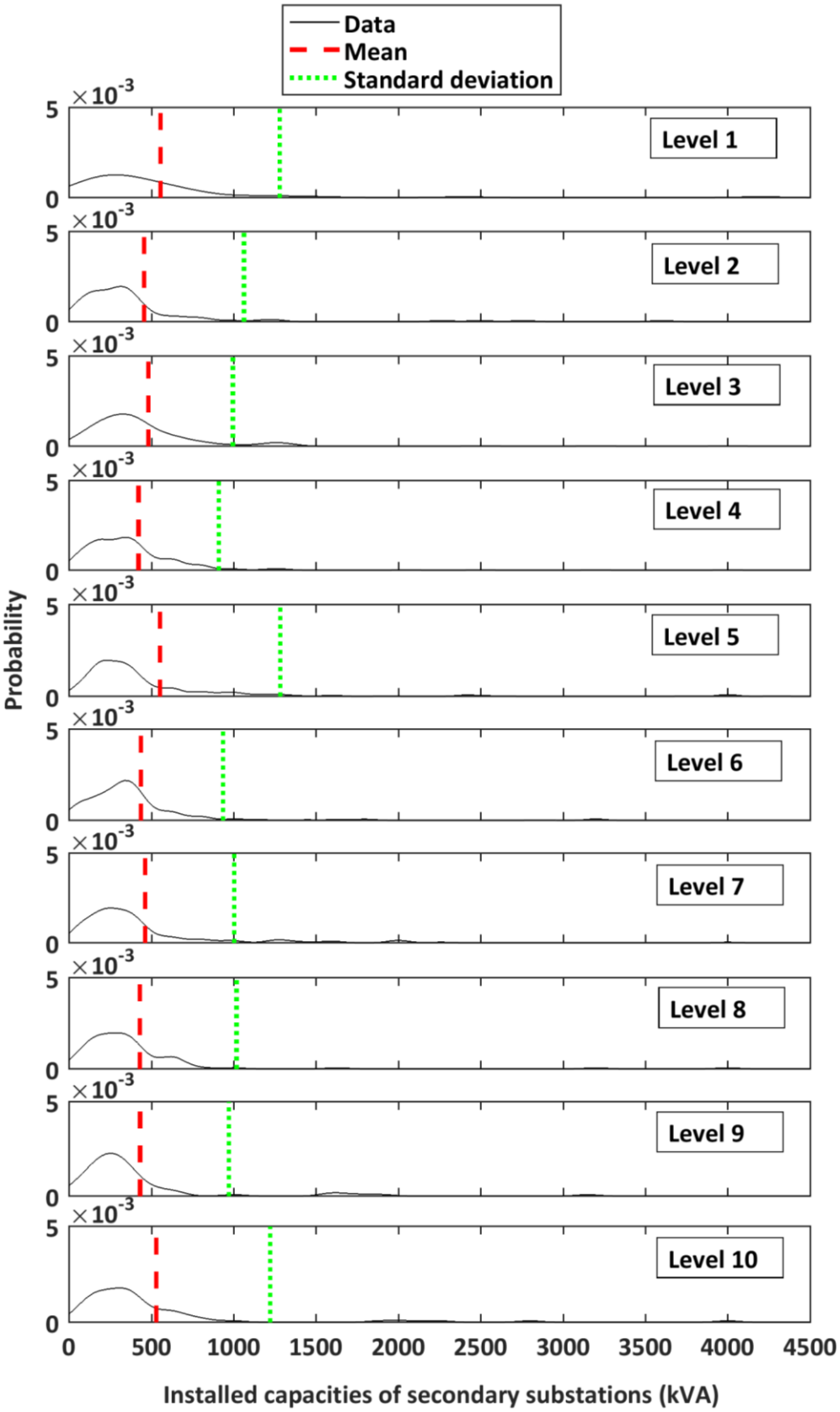


Figure 4.10: Level dependent distribution of the installed capacities of secondary substations in urban networks.

4.8 Summary of the investigation of electrical properties

This chapter studied the electrical properties of real-world electricity distribution networks at the MV level. A limited set of available data regarding the installed capacities of distribution substations and the conductor cross sections of the distribution lines were used for the study. The sub-urban and urban grouping of the networks is validated using a clustering approach. A novel approach to obtain depth dependent electrical properties has also been developed. Kernel density PDFs for the secondary substation capacities with the distance from the source nodes of the two types of networks were investigated.

Results from the real-world network investigation showed that,

(i) the substations capacities and the conductor cross sections are able to characterise the electrical features of sub-urban and urban networks.

(ii) PDFs which describe the distribution of secondary substation capacities along the feeder lengths of the two types of networks have clear differences.

Chapter 5: Development of the statistically-similar networks generator

5.1 Introduction

The detailed development process of the statistically-similar electricity networks generator (SSNG) proposed in this thesis is presented in this chapter (Figure 1.3). The key feature of the SSNG is its ability to generate large numbers of random, realistic models of electricity distribution networks which are statistically-similar in terms of a set of topological and electrical properties as defined by the user with some given values or ranges of values

The overview of the development process of the SSNG is shown in Figure 5.1. The SSNG is a data driven model. The development process of the SSNG consists of four stages: data collection and preparation, feature identification, network generation, and validation. These four stages of the development process of the SSNG are independent from the chosen network data set. However, the detailed approach in each stage is varied according to the details of the available data. This chapter describes the development of the SSNG using a set of 10kV level distribution network data collected from the Chinese power grids.

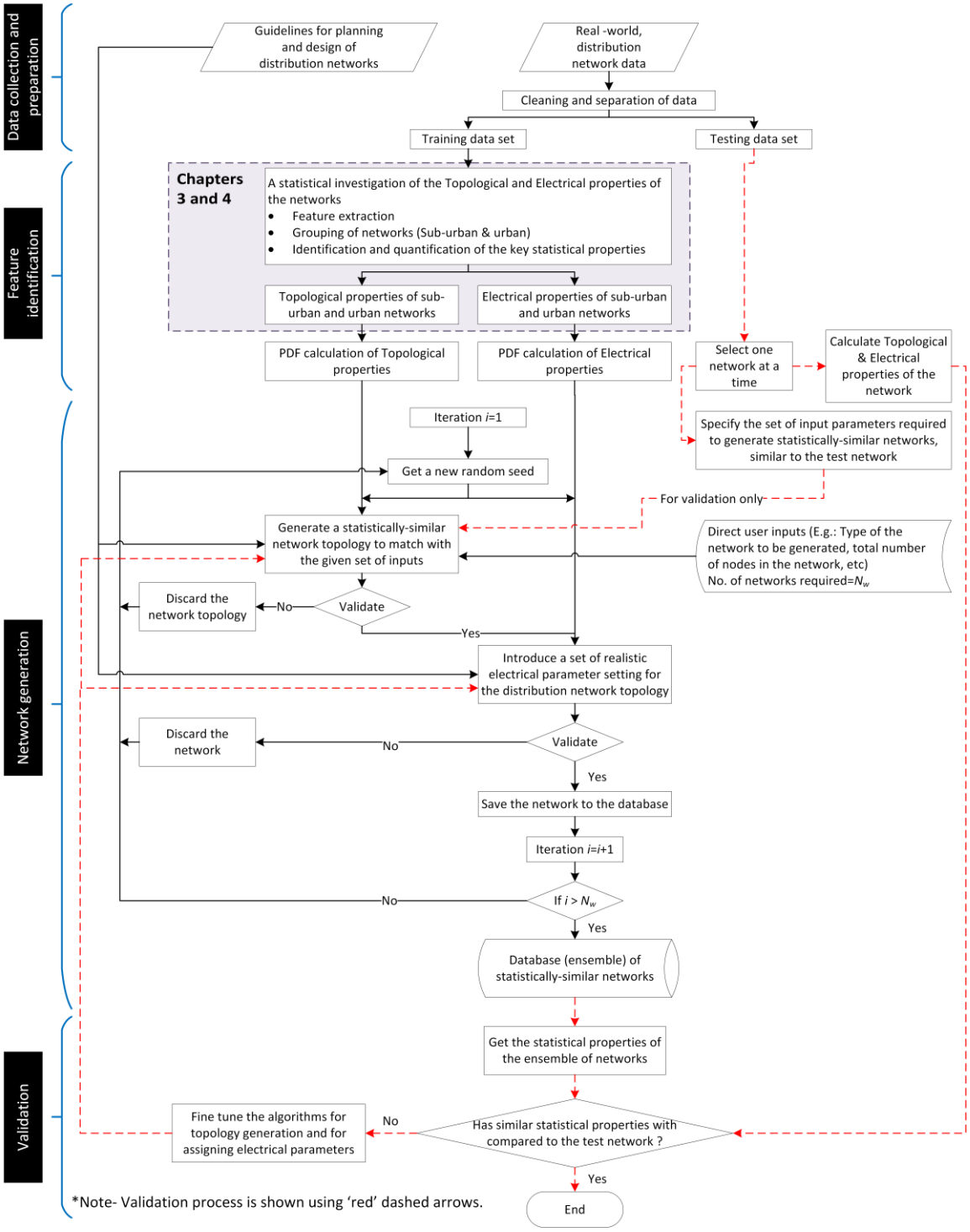


Figure 5.1: Overview of the development procedure of the SSNG.

- **Data collection and preparation stage**

There are two key inputs to the SSNG: real-world distribution network data and guidelines for planning and design of electricity distribution networks. A set of real-world distribution

networks were collected. Then, the planning and design guidelines which are applicable for the selected set of real-world networks were identified. The selected set of real-world networks was then divided in to two datasets, namely *training* and *testing* data set. The network data in the *training* dataset was used in the training process where, the different network types and various statistical properties of different network types were identified and quantified. The *testing* dataset represents the real network data and was used to validate generated network models against the real networks. The largest part of the data formed the *training* dataset whereas the rest were used in the *testing* dataset.

- **Feature identification stage**

Identification and quantification of the important statistical (topological and electrical) properties of real-world distribution networks is a key requirement for developing random, but realistic network models. Therefore, an investigation of the statistical properties of real-world distribution networks was conducted using the training data set. This step of the development process of the SSNG is described in Chapters 3 and 4 of this thesis. Probability density functions (PDFs) of topological and electrical properties in different types of distribution networks (urban and sub-urban) were obtained from this investigation.

- **Network generation stage**

The PDFs of topological and electrical properties of real-world networks, together with the relevant guidelines for network planning and design were then used to develop the statistically-similar network generator. The network generator was developed in a hierarchical manner. One iteration of the algorithm (E.g.: *iteration i* =1) is explained here as an example. First, a realistic model of the network topology is generated by taking into account the user inputs regarding the networks to be generated. Then, the realistic electrical parameter settings are assigned to the above network topology model. There are two separate validation procedures within the

hierarchical approach. Generated network topologies and the assigned electrical parameter settings of the network topologies are validated against the network planning and design guidelines. The outcomes (i.e. topologies and topologies assigned with electrical parameter settings) that violate the design standards are discarded.

Likewise, starting with different random seeds the network generator is able to generate as many statistically-similar distribution networks as specified by the user.

- **Validation stage**

A validation stage was used to validate the networks generated by the SSNG against the real networks. An ensemble of statistically-similar networks to a selected network from the *testing* dataset was generated by the SSNG. Then, the statistical properties of the ensemble of the statistically-similar networks are compared with that of the selected network from the *testing* dataset.

The detailed network generation procedure and the validation of the generated network models against *testing* data set are presented in the following sections.

5.2 Guidelines for network planning and design

This section provides a summary of the guidelines for network planning and design, used during the statistically similar networks generation. Identifying the relevant guidelines for network planning and design is important when generating realistic networks in order to ensure that the networks are consistent with the planning and design standards. In the development process of the SSNG, guidelines for network planning and design were used to achieve the following goals.

(i) To compare and validate the results of the statistical investigation of the real-world distribution networks with the guidelines for network planning and design. A successful

validation against the planning and design guidelines gives the confidence that the results of the statistical investigation reflect both the characteristics of the real-world networks and the planning and design standards.

(ii) To ensure that the networks generated by the SSNG do not violate the planning and design standards of the distribution networks.

The guidelines for network planning and design are usually different from country to country, region to region and also from the distribution network operator (DNO) to operator. The supporting network data for the development of the SSNG is collected from Chinese distribution networks. Hence, this study refers to the guide for planning and design of the Chinese distribution networks [75].

5.2.1 Guidelines for planning and design of Chinese distribution networks

In the planning process of Chinese electrical power networks, the power supply areas are divided into different zones according to their load densities σ, which is measured in MW per km^2 (Table 5.1). According to the guidelines, in order to obtain σ the minimum area must be above 5 km^2.

Table 5.1: Division of power supply zones [75].

Power supply zone		A+	A	B	C	D	E
Administrative level	**Directly under the central government**	City centre area $\sigma \geq 30$	Downtown $15 \leq \sigma < 30$	Downtown $6 \leq \sigma < 15$	Urban area $1 \leq \sigma < 6$	Sub-urban area & country side $0.1 \leq \sigma < 1$	-
	Provincial capital	City centre area $\sigma \geq 30$	City centre area $15 \leq \sigma < 30$	Downtown $6 \leq \sigma < 15$	Urban area $1 \leq \sigma < 6$	Sub-urban area & country side $0.1 \leq \sigma < 1$	-
	Prefecture level city	-	City centre area	City centre area	Downtown & urban area	Sub-urban area & country	unpopulated areas

		$\sigma \geq 15$	$6 \leq \sigma < 15$	$1 \leq \sigma < 6$	side $0.1 \leq \sigma < 1$
County level city	-	-	urban area $\sigma \geq 6$	urban area $1 \leq \sigma < 6$	Sub-urban area & country side $0.1 \leq \sigma < 1$

Depending on the power supply zone, planning guidelines for 10kV distribution networks such as selection of the network structure (E.g.: radial, single/double ring), capacities of the main supply transformers and selection of conductors (OH and UG) are then decided (Table 5.2). Recommendations are provided to choose the conductor type and the cross sections for 10kV distribution lines depending on the main transformer capacities (Table 5.3). The recommended number of (10 kV) outgoing feeders from the main transformer depending on its capacity is also listed in the guidelines and shown in Table 5.3. Moreover, some recommendations regarding the installed capacities of the pole mounted 10kV secondary substations are shown in Table 5.4.

Table 5.2: Planning standards of distribution networks according to the power supply zones [75].

Power supply zone	Sub-station transformer capacity	10 kV line type	Structure of the network
A+, A	High/medium	Underground cables must be used. Overhead cables are used occasionally	UG cables: double ring, single loop, n-1 ($2 \leq n \leq 4$) OH lines: Multi sectional
B	High/medium	Overhead lines are used. Underground cables can be used at an additional cost.	UG cables: single loop, n-1 ($2 \leq n \leq 4$) OH lines: Multi sectional
C	Medium/small		UG cables: single loop OH lines: Multi sectional
D	Small	Overhead lines are widely used.	OH lines: Multi sectional, radial
E	Small	Overhead lines are widely used.	OH lines: radial

Table 5.3: 10 kV distribution line information/recommendations depending on the main transformer capacity [75].

Main transformer capacity (MVA)	Number of 10kV feeders	Cross section area of 10 kV trunk lines/main feeders (mm^2)		Cross section area of 10 kV lateral branches (mm^2)	
		OH lines	UG cables	OH lines	UG cables
80, 63	≥ 12	240, 185	400, 300	150, 120	240, 185
50, 40	8 ~ 14	240, 185, 150	400, 300, 240	150, 120, 95	240, 185, 150
31.5	8 ~ 12	185, 150	300, 240	120, 95	185, 150
20	6 ~ 8	150, 120	240, 185	95, 70	150, 120
12.5, 10, 6.3	4 ~ 8	150, 120, 95	-	95, 70, 50	-
3.15, 2	4 ~ 8	95, 70	-	50	-

Table 5.4: Recommended capacities for 10kV pole mounted transformers [75].

Power supply zone	3 phase pole mounted transformer capacity (kVA)
A+, A, B and C	≤ 400
D	≤ 315
E	≤ 100

In addition to the guidelines given in the above tables, there are several other rules used in the planning and design of Chinese distribution networks;

- It is recommended to use a maximum of three different cross sections in one distribution feeder.
- For A+, A and B type power supply areas, the supply radius from the main transformer should be within 3 km.
- For type C supply areas, the maximum supply radius should be ≤5 km.
- For type D supply areas, the maximum supply radius should be ≤15 km.

5.2.2 Comparison of the statistical properties of real-world networks with the guidelines for network planning and design

It was identified in Section 5.2.1 that the guidelines for planning and design of the distribution networks are structured according to the power supply zones which are defined by the load densities σ (measured in MW/km^2). However, the load density information was not available in the collected set of real-world network samples. In Chapter 4, to overcome this limitation of data, approximate values for the load density of the networks (LD), were derived using the installed capacities of the secondary substations (Equation 4.10). The units of LD are in MVA/km^2. Therefore, an approximate value for the load densities in real world networks, σ_{approx} in MW/km^2 can be obtained as shown in Equation (5.1).It is assumed that pf is the power factor at which each secondary substation in the network is operating under normal conditions.

$$\sigma_{approx}(\mathrm{MW/km^2}) = LD(\mathrm{MVA/km^2}) \times pf \tag{5.1}$$

These values obtained for σ_{approx} in the real-world networks, enable the comparison of the statistical properties of real world electricity distribution networks with their planning and design guidelines as shown in Table 5.5. The outcomes of the comparison are summarised below.

- According to the ranges of variation of the σ_{approx} of the real-world sub-urban and urban networks derived using the data in Table 4.5 and the Equation (5.1), the urban group of networks corresponds to the power supply zone C and sub-urban group of the networks corresponds to the power supply zone D.

Table 5.5: Comparison of the statistical properties of real-world networks with the guidelines for network planning and design.

Properties of real-world networks from the investigations in Chapters 3 and 4			**Planning and design guidelines**		
Property	**Urban networks**	**Sub-urban networks**	**Property**	**Type C supply areas**	**Type D supply areas**
$\boldsymbol{\sigma_{approx}}$ (MW/km2)	$1.31 \leq \sigma \leq 4.4$	$0.1 \leq \sigma \leq 0.21$	σ (MW/km2)	$1 \leq \sigma \leq 6$	$0.1 \leq \sigma \leq 1$
Capacities of main transformers used in the primary substations (MVA)	16, 20,31.5, 50, 63	5, 6.3, 8, 10, 16, 20	Capacities of main transformers used in the primary substations (MVA)	Medium-small (E.g. 16, 20,31.5, 50, 63)	Small (E.g. 5, 6.3, 8, 10, 16, 20)
Number of outgoing feeders from primary substations	6-10	4-7	Recommended number of outgoing feeders from main transformer	6-14	4-8
Maximum recorded feeder lengths (km)	5.6	15.1	Maximum supply radius constraint (km)	≤ 5	≤ 15
The choice of cable cross sections (mm^2)	95,120,185,240,300	35,50,70,95,120	Recommendations for the choice of cable cross sections (mm^2)	95,120,185,240, 300,400	35,50,70,95, 120
Average installed capacity of a secondary substation (kVA)	367 - 731	145 - 258	Recommended capacities of pole mounted 10kV transformers (kVA)	≤ 400	≤ 315

- According to the guidelines in Table 5.3, there are no hard rules of defining large, medium and small capacities of the main transformers used for the primary substations. Therefore, with the assumptions that medium to small capacities are in the range of 2 MVA- 65MVA, it is reasonable to say that the capacities of the main transformers of the urban and sub urban networks match well with the recommendations for power supply zones C and D respectively.
- The number of outgoing feeders from the main transformers in the urban and sub-urban networks are also consistent with the standards of network planning and design in Table 5.3.
- Maximum supply radius constraints from the planning guidelines are compared with the maximum feeder lengths of the real networks. Even though the definitions of these two measures are not the same, they both give an idea about the distance to the furthest node in the network from the supply node. In real world sub-urban networks, maximum recorded feeder length of 15.1km is just above the maximum supply radius constraint of type D supply zone which is 15km. The maximum recorder feeder length of urban networks has a slight deviation of +12% from the power supply zone C constraint for maximum supply radius.
- The choice of cable cross sections used in the sub-urban and urban networks are similar to the conductor cross-sections specified by the network planning guidelines in Table 5.3.

The outcomes of the above comparison prove that the selected set of real-world networks are consistent with the network design standards and also validates the statistical investigation in Chapters 3 and 4.

5.3 Generation of statistically-similar network topologies

The generation of statistically-similar networks begins with the network topology generation. Starting with different random seeds and a given set of input parameters, the idea is to generate different network layouts (topologies) which have similar topological properties (Figure 5.1).

The networks are generated as graph models and visualised in the form of tree diagrams to show the connection between the nodes. The proposed SSNG does not generate actual graphical portrayals of the network topologies (i.e. visualisations in terms of actual lengths and the realistic angle distributions of the network branching are not generated).

5.3.1 The inputs required by the algorithm

The input parameters required by the algorithm can be categorised into three groups; direct user inputs, inputs derived from the topological investigation of real-world networks and the constraints from the guidelines for network planning and design.

- **Direct user inputs**

It is expected that several characteristics of the networks to be generated is provided by the user of the network generating tool. These are defined as direct user inputs. Some of these direct user inputs can be derived from the available set of information of the network to be analysed.

(i) Network type (urban/sub-urban)

If the total population and the power supply area is known, according to the population density the type of the network (urban/sub-urban) can be identified. Load density parameter can also be used to identify the network type according to Table 5.1.

(ii) The total number of nodes/ number of secondary substations

Two options have been given to input the number of nodes in the network. The user must input either the total number of nodes *N*, or the number of secondary substations in the network.

In addition,

(iii) Capacity of the main supply transformer/primary substation and

(iv) The number of outgoing feeders from the main transformer/primary substation

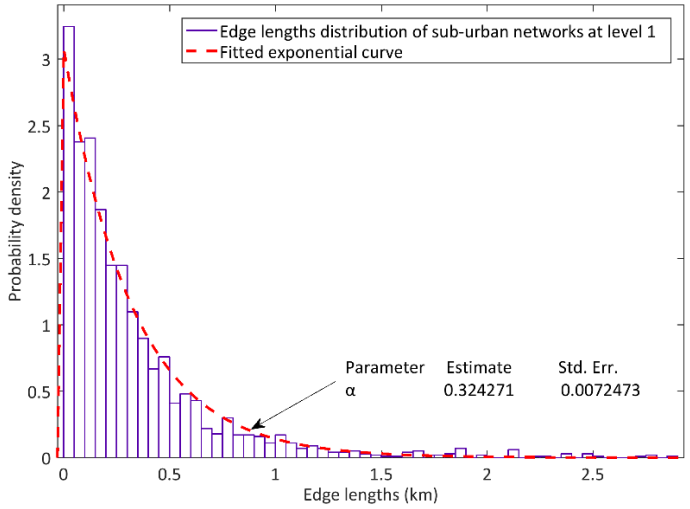
are required from the user as inputs to the topology generating algorithm

- **Inputs generated from the topological investigation of real-world networks**

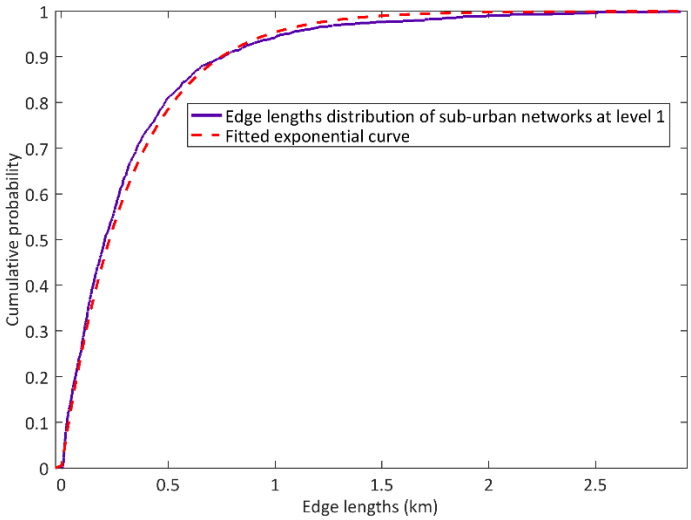
From the topological investigation in Chapter 3, three sets of inputs are required to the network topology generation. These inputs must be in the form of probability density functions (PDFs) and probability mass functions (PMFs). A PMF differs from a PDF in that the latter is associated with continuous random variables rather than discrete random variables [78].

(i) Depth dependent edge length distributions of urban and sub-urban networks

The depth dependent edge length distributions are represented in the form of PDFs, as the value of an edge length can be any non-integer value between the specified limits (i.e. edge length is a continuous variable). Therefore, the resulting PDFs of the depth dependent edge length distributions are approximated by well-known distributions (E.g. exponential distribution, lognormal distribution, gamma distribution, etc.). It was identified through distribution fitting exercises that, negative exponential distribution curves show the best fit with the PDFs of depth dependent edge length distributions. Figures 5.3(a) and 5.3(b) show the PDF and the cumulative density function (CDF) of the edge length distribution of all the sub-urban networks at 'level 1'. The PDF and CDF of the best fit exponential distribution are also plotted on top of the same figures.



(a)



(b)

Figure 5.2: (a) PDF and (b) CDF of the edge length distribution of the sub-urban networks at 'level 1' of a network and the fitted exponential distribution on the same figure.

According to Figure 5.2, the best fit exponential distribution can be represented using the mean of the exponential distribution (α=0.324) and its standard error (*Std. Err*. = 0.00724) compared to the mean of the actual dataset. The corresponding parameters for the edge length distributions of all the levels of urban and sub-urban networks are listed in Table 5.6. Therefore, a general representation for the PDFs of the depth dependent edge length distributions of the sub-urban and urban networks (at each level separately), $f(l, \alpha)$ can be obtained as shown in the Equations (5.2) and (5.3). The $l_{\min}$ and $l_{\max}$ refer to the minimum and maximum edge lengths observed at each level of the two types of real-world networks.

$$f(l,\alpha) = \begin{cases} 0\,; & l < l_{\min} \\ \frac{1}{\alpha} e^{-\frac{l}{\alpha}}; & l_{\min} \le l \le l_{\max} \\ 0; & l > l_{\max} \end{cases} \quad (5.2)$$

$$\int_0^{\infty} f(l,\alpha)\, dl \approx 1 \quad (5.3)$$

Table 5.6: Parameters of the PDFs of depth dependent edge length distributions of sub-urban and urban networks.

Level of the network, i	Sub-urban networks, t=0				Urban networks, t=1			
	α	*Std.Err.*	$l_{\min}$ (km)	l_{max} (km)	α	*Std.Err.*	$l_{\min}$ (km)	$l_{\max}$ (km)
i = 1	0.324	0.00724	0.02	2.9	0.149	0.00384	0.01	1.4
i = 2	0.264	0.00630	0.02	2.0	0.108	0.00323	0.01	0.9
i = 3	0.236	0.00853	0.02	1.6	0.087	0.00402	0.01	0.7

(ii) Depth dependent degree distributions of urban and sub-urban networks

The depth dependent degree distributions are expressed in terms of PMFs as the degree of a node can only be a positive integer value (i.e. the degree of a node is a discrete variable). Equation (5.4) shows a general expression for the depth dependent degree distributions of sub urban and urban networks at the levels i =1, 2, 3. The parameter k stands for the degree of a node in a network and k_{max} stands for the maximum node degree observed in a network. $f_i(k)$ stands for the probability that a degree of a node in the network level i is equal to k. For example, p_{i1} is the probability that a degree of a node in the network level i is equal to 1.

$$f_i(k) = \begin{cases} 0; & k \le 0 \\ p_{i1}; & k = 1 \\ p_{i2}; & k = 2 \\ \quad\vdots & \quad\vdots \\ p_{in}; & k = k_{max} \\ 0; & k > k_{max} \end{cases} \quad (5.4)$$

$$\sum_{k=1}^{k_{max}} p_{ik} = 1 \quad (5.5)$$

(iii) Distribution of the nodes among the levels of urban and sub-urban networks

If the total number of nodes in a network is *N*, and the probability that a given node is in level 1, 2 or 3 are respectively $q1$, $q2$ and $q3$, then the PMF for the distribution of the nodes among the levels of a network $f(n)$, can be expressed as shown in the Equations (5.6) and (5.7). (i.e. the level of a node in a network is a discrete random variable and hence $f(n)$ is in the form of a PMF).

$$f(n) = \begin{cases} q_1; & level = 1 \\ q_2; & level = 2 \\ q_3; & level = 3 \end{cases} \tag{5.6}$$

$$q_1 + q_2 + q_3 = 1 \tag{5.7}$$

As mentioned in the 'direct user inputs' section there are two options to input the number of nodes in the network to the network generation tool. This is because, it may be difficult for the user to know the total number of nodes in a network including the busbars, switching locations and substations, etc. Getting a number for the total number of secondary substations in the network may be practically easier than getting the total number of nodes in the network. Therefore, if the number of secondary substations (number of leaf nodes, n_{leaf}) in a network is given, now the total number of nodes, *N* in a network can be derived using the above probabilities, as shown in the Equation (5.8).

$$n_{leaf} = N(\, q_1 \times p_{11} + q_2 \times p_{21} + q_3 \times p_{31}) \tag{5.8}$$

Equation (5.8) can be explained as follows. If the total number of nodes in the network is *N*, the total number of nodes in the level 1 of the network is equal to $N \times q_1$. Similarly, the total number of nodes in the levels 2 and 3 become $N \times q_2$ and $N \times q_3$ respectively. The number of leaf nodes in one level of the network can be obtained by multiplying total number of nodes in that level with the probability of a node in that level having a degree equals to 1 (E.g.: The number of leaf nodes in level 1= $N \times q_1 \times p_{11}$). The addition of the leaf nodes in all the three levels is equal to the total

number of leaf nodes in the network. This will lead to the Equation (5.8) and *N* can be calculated by substituting the values for probabilities and the number of leaf nodes to the equation.

All the parameters introduced in the above equations to describe the PDFs and PMFs of topological properties of real-world networks are called the 'SSNG model parameters'.

- **Inputs from the guidelines for distribution network planning and design**

All the inputs from guidelines for network planning and design are used as the constraints (i.e. rules to follow) in the algorithm for the network topology generation. This is to ensure that the generated topologies do not violate the standards of actual network design.

(i) Guidelines for the capacities of the main supply transformers (i.e. depending on the network type) and

(ii) Guidelines for the number of outgoing feeders from the main transformer depending on its capacity are used as initial checks for the direct user inputs.

(iii) The guidelines for the maximum supply radius (r_{supply_max}) of urban and sub-urban networks are used at the end of each topology generation, to discard the networks that exceed these supply radius limits.

5.3.2 The algorithm for the generation of statistically-similar network topologies

The flow chart of the algorithm for generation of network topologies is shown in Figure 5.3. The algorithm explains the generation of a single network topology. The process shown in the algorithm is repeated to generate many statistically-similar network topologies. Each step of the algorithm is explained in this section. The main steps are numbered in Figure 5.3 and the step-wise explanation of the algorithm follows this numbering.

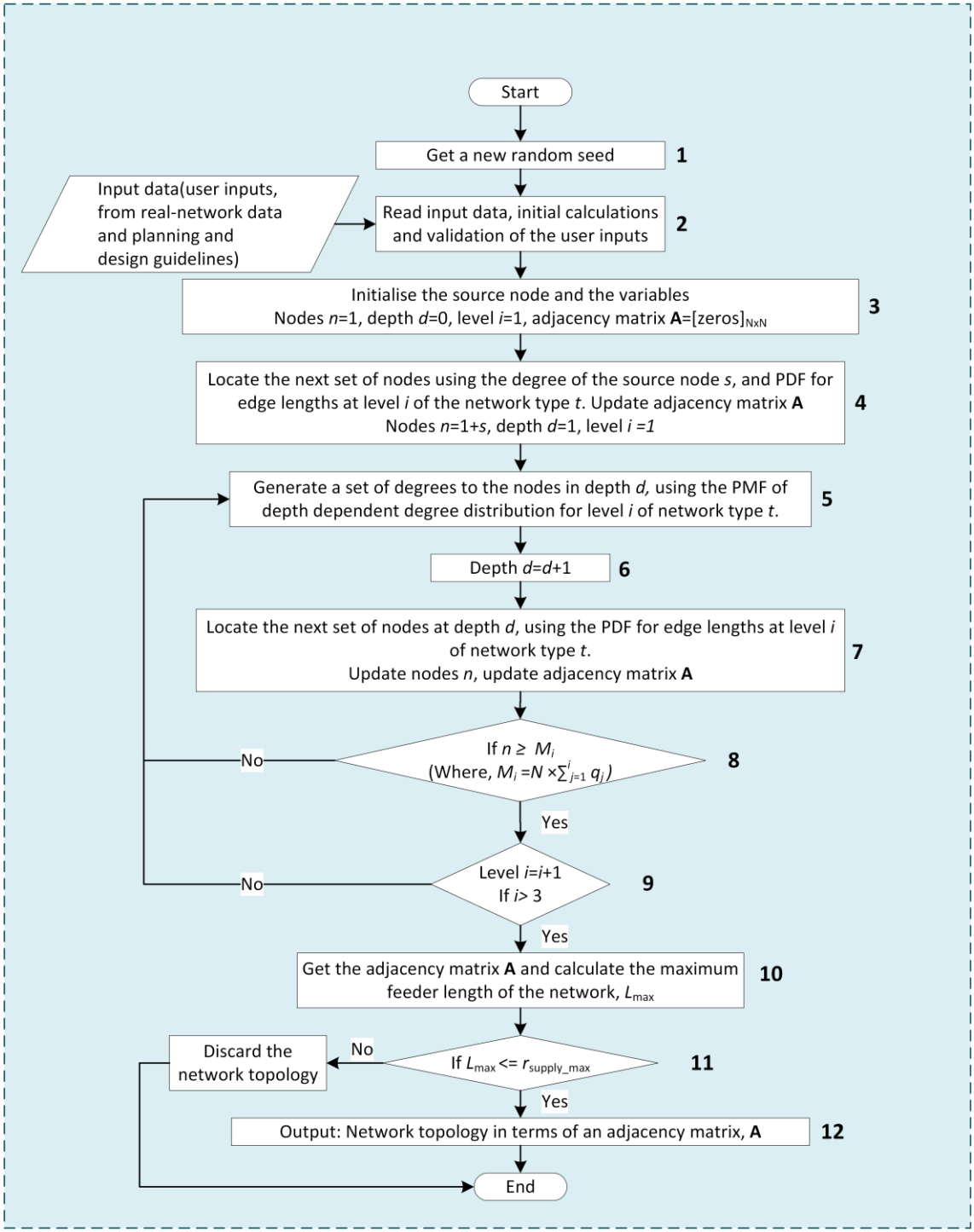


Figure 5.3: Flowchart of the algorithm for the generation of network topologies.

The topologies are generated as evolving network topologies. Starting with one source node (to represent the primary substation of the radial network), gradually new nodes are added and connected to the existing network topology. The PDFs of edge length and the PMF's of node degree are used to reproduce realistic edge lengths and branching of the networks generated.

It is assumed that the edge length, degree of a node and the level of a node in a network are statistically independent events, which means the occurrence of one event does not affect the probability of occurrence of the other [79].

1. A new iteration of the topology generation starts with a new random seed. This is to ensure that, at each iteration of the algorithm a different set of values are produced from the probability distributions of edge lengths and degree of the nodes in the steps described below.
2. Then the direct user inputs are checked against the constraints of the network planning and design guidelines. Type of the networks, t to be generated is identified from the user inputs. The calculation of the variable M_i, for the levels of the network i =1,2,3 is important in the generation process of the network topology. M_i for is calculated according to Equation (5.9).

$$M_i = N \times \sum_{j=1}^{i} q_j \tag{5.9}$$

   M_1 represents the total number of nodes in the level 1 of the network, M_2 represents the total number of nodes in the levels 1 and 2 of the network and similarly M_3 represents the total number of nodes in all the three levels of the network which is also equal to *N*. q_j for *j*=1,2,3 is the probability that a given node in the network belongs to the level *j* of the network. M_i is used to decide the change of the network level according to the number of nodes in the evolving network topology.

3. The source node of the network topology is initialised at first. Depth of the network at the source node is equal to zero (*d*=0). The number of nodes in the evolving network topology is denoted by the variable *n* and is equal to 1 at this step of the algorithm. The

source node belongs to the level 1 (i = 1) of the network. Initialising the adjacency matrix, $\boldsymbol{A}$ is important in this step. The output of the algorithm is a network topology which is given in terms of a weighted adjacency matrix as shown in Equation (5.10). $\boldsymbol{A}$ is a *N*×*N* square matrix. If there is an edge between nodes *x* and y, that are added to the network topology, the element (x, y)) equals to the length of the edge ($l_{x,y}$), between those two nodes. Otherwise $\boldsymbol{A}$ (x, y) is equal to zero. At this stage there are no edges in the network, only the source node exists. Therefore, initial $\boldsymbol{A}$ is a *N*×*N* square matrix with all zeros.

$$\boldsymbol{A}(x,y) = \begin{cases} l_{x,y}, & if\ there\ is\ an\ edge\ between\ nodes\ x, y \\ 0, & Otherwise \end{cases} \tag{5.10}$$

4. The number of nodes at the next depth level (at *d*=1) is decided by the degree of the source node, k_s as specified by the user. Length of the edges which connect the source node with the nodes at the depth *d*=1, are assigned with random length values which are generated from the PDF of the edge length distribution of network type *t*, at level $i = 1$. This step of the algorithm is illustrated in Figure 5.4 for a case where, k_s= 3.

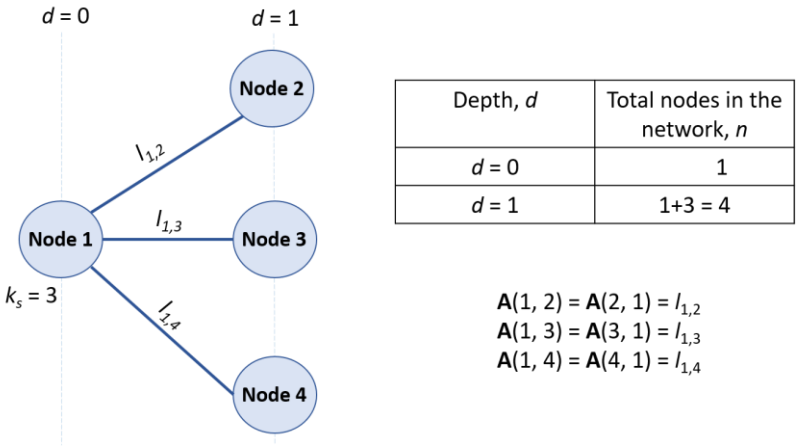


Figure 5.4: Development of the network topology up to depth, *d*=1.

All the variables are then updated. The total number of nodes *n*, of the evolving network topology is now equal to 1+k_s. The network level, i is still equal to 1. Depth *d* of the

network is also equal to 1. The elements in $\boldsymbol{A}$ are updated according to the connections between the nodes. An example is shown in Figure 5.4. (Note: The actual angles between the links are ignored in the network generation process).

5. Then a set of random values are assigned to the degrees of the nodes at the depth, *d*=1 of the network. These random values for the node degrees are generated using the PMF of the node degree distribution for the level, $i = 1$ of the network type, *t*.

6. After assigning the degree to the nodes at depth, *d* = 1 the network can now evolve to the next depth, *d* = *d*+1.

7. The next step is similar to the fourth step described above. The next set of nodes at the depth, *d* = 2 of the network are assigned using the PDF for the edge lengths distribution at level, $i = 1$ of the network type, *t*. Again, all the variables are updated. This step of the algorithm is illustrated in Figure 5.5. It was assumed that the random degrees of the nodes 2, 3 and 4, generated by the PMF of the degree distribution (in the step 5) are equal to 3, 3 and 2 respectively.

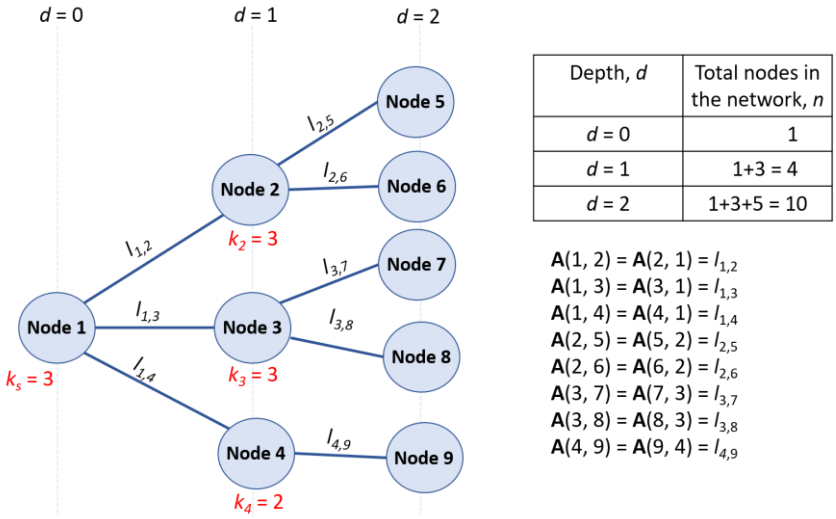


Figure 5.5: Development of the network topology up to depth, *d*=2.

8. Steps 5 to 7 are repeated until the number of nodes, *n* reaches the value M_1 which is the total number of nodes in the level 1 of the network.

9. If the number of nodes, *n* exceeds M_1, the variable i which denotes the level of the network is updated to the next level, $i = i + 1$. The algorithm repeats Steps 5 to 7 with the PDFs of edge lengths and PMFs of node degrees, for the new level ($i = 2$) of the network in order to locate the new nodes which are joining the network. The aim of this process is to use the topological characteristics of real-world networks at different 'levels' to generate the topological structures of artificial networks. Therefore, the algorithm first completes the generation of 'level 1' of the network topology and progressively moves to develop the next two levels of the network topology. If the number of nodes, *n* reaches the value M_3 (i.e. M_3 is equal to *N*) and i exceeds the number of levels of the network (i > 3) the algorithm stops adding more nodes to the existing topology and exits from all the loops.

10. The resulting network topology can now be studied using the weighted adjacency matrix, $\boldsymbol{A}$ for its topological properties. The maximum feeder length, L_{max} of a network

topology is the length of the path (i.e. is the addition of the lengths of all the edges in the path) from source node to the furthest away node from the source node.

11. The maximum feeder length, L_{max} is compared with the the maximum supply radius, r_{supply_max} in the planning and design guidelines. If $L_{max} > r_{supply\_max}$ the network topology generated by the algorithm violates the network design standards. Then the topology will be discarded.

12. If the network topology generated by the algorithm complies with the above design standard, it will be saved and forwarded to the next algorithm (Section 5.4) to assign the electrical parameter settings to the network topology.

5.4 Generating realistic, electrical parameters for the network topologies

In this part of the study, it is assumed that the topological distribution of consumers is captured by the models of different types of electricity distribution networks (urban and sub-urban). Then, the electrical parameters are assigned to the network topologies using the results of statistical analysis of electrical properties of the real-world networks and the guidelines for network planning and design. The detailed process is described in this section.

5.4.1 The inputs required by the algorithm

The inputs required to assign electrical parameters to the distribution network topologies come from four different sources: directly from the user, as the outputs from the topology generating algorithm, from the investigation of electrical properties of the real-world networks and also from the guidelines for network planning and design.

- **Inputs from the user**

These user inputs are first used in the topology generating algorithm and passed down to the algorithm for electrical property assignment (as shown in Figure 5.1).

(i) Network type, *t* is important to identify the correct set of Kernel density probability distributions of the secondary substation capacities derived from the real-world network data.

(ii) The capacity of the main supply transformer is used for the selection of conductor cross sections for the distribution lines according to the network planning guidelines.

- **Input that comes as an output from the topology generating algorithm**

Network topology is the most important input for this algorithm. It is specified in terms of a weighted adjacency matrix, $\boldsymbol{A}$ and is generated by the topology generating algorithm described in Section 5.3.

- **Inputs from the investigation of electrical properties of the real-world networks.**

An investigation of the electrical properties of real-world networks was carried out in Chapter 4. Three set of inputs are taken from results of this investigation.

(i) The Kernel density probability distributions of the secondary substation capacities along the feeder lengths of the urban and sub urban networks.

These Kernel density probability distributions, are used to generate a realistic set of secondary substation capacities to a given network topology. However, the values

generated from a continuous PDF can be any non-integer value. Therefore, these values must be approximated to the nearest value of actual secondary substations.

(ii) The list of secondary substation capacities observed in the real networks.

The list of actual secondary substation capacities is used to approximate the random capacities generated by the above kernel density probability distributions to the nearest value of actual secondary substations.

(iii) Electrical properties of the conductors used in the distribution lines of the real-world networks

Electrical properties of the conductors (E.g. per-km resistance, per-km reactance, kVA rating) used in the distribution lines of the real-world networks are given to the algorithm as another set of inputs. These properties are obtained according to the conductor type and the cross section (Table 4.3).

- **Inputs from the guidelines for distribution network planning and design**

(i) A selection of conductor cross sections for the distribution lines depending on the main transformer capacities are given as a set of inputs to the algorithm from the guidelines for network planning and design.

(ii) A constraint for the maximum allowable voltage drop of the networks is used to ensure that the generated networks are not violating the design standards.

5.4.2 The algorithm for assigning realistic electrical parameters for the network topologies

The flow chart of the algorithm for assigning realistic electrical parameters for distribution network topologies is shown in Figure 5.6. Each step of the algorithm is described as shown by the number in the figure.

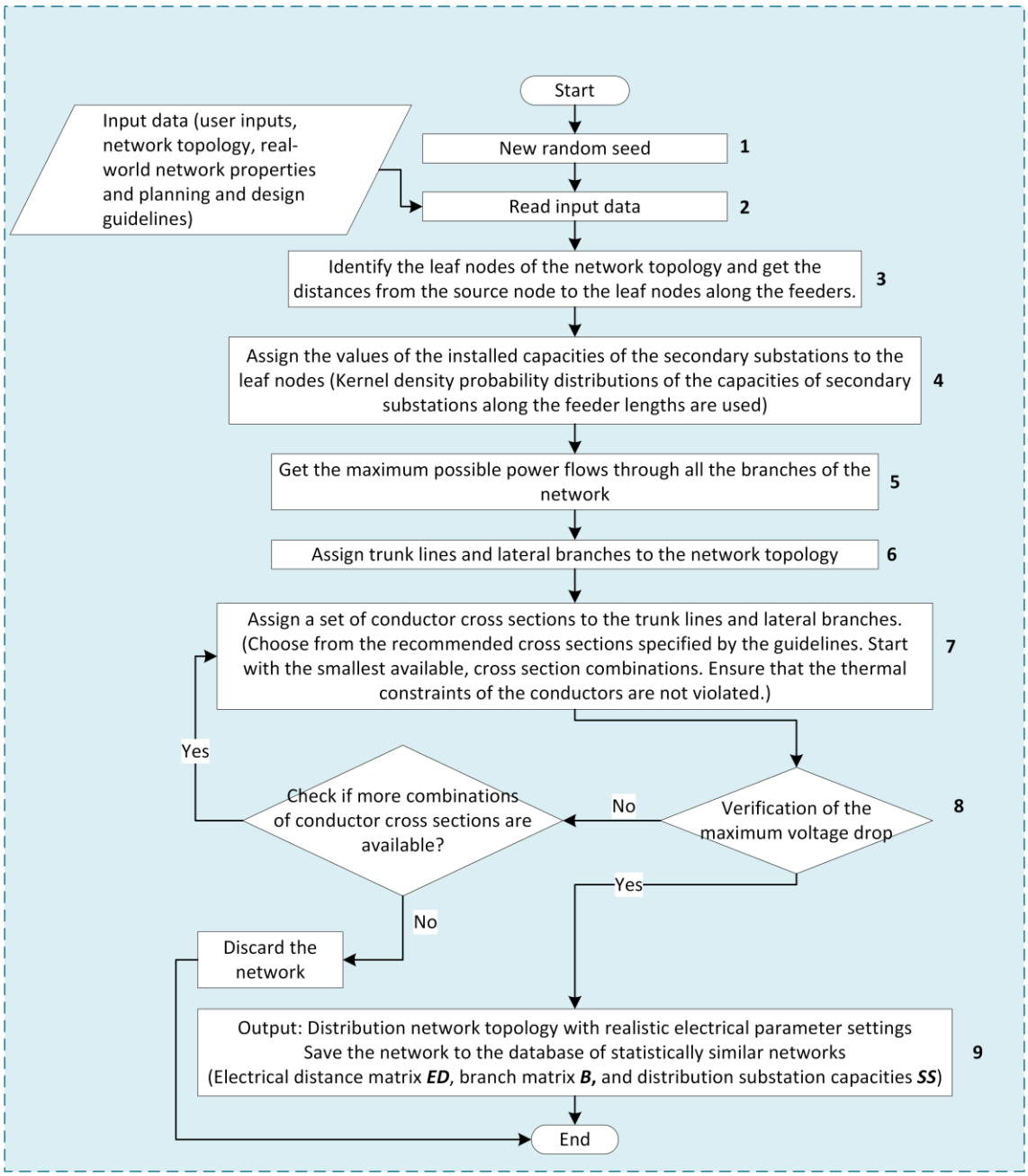


Figure 5.6: Flowchart of the algorithm for the assignment of realistic electrical parameters for network topologies.

1. Similar to the topology generating algorithm, each iteration of this algorithm starts with a new random seed. In fact, in one iteration of the complete network generation process, (i.e. hierarchical approach in Figure 5.1) the algorithm for topology generation and assignment of electrical parameters share the same random seed. Different random seeds ensure that different set of values are produced by the PDFs in the different iterations of the algorithm.

2. The next step is to read all the input parameters given to the algorithm. The type, *t* of the network to be generated is identified.

3. According to the study of real-world networks in Chapter 4, leaf nodes of a network topology represent the secondary substations (Figure 4.4). Therefore, it is important to identify the leaf nodes (i.e. nodes with a degree equals to 1) in the network topology. Next, the distance from the source node to each of the leaf node is calculated. Djakarta's shortest path algorithm is used to obtain these distances [65]. These distances will help in the next step to assign the substation capacities to the leaf nodes.

4. Assuming a uniform distribution of the consumer load along the feeders of the network topology is not realistic approach. Therefore, results from the statistical investigation of electrical properties in Chapter 4 was used to allocate the capacities of secondary substations for all the leaf nodes in the network topology to obtain an approximation to the actual consumer load distribution. The secondary substation capacities were assigned depending on the distance to the source node. These capacities are generated by the corresponding Kernel density probability distributions and approximated to the nearest values of actual capacities of secondary substations in Table 4.2.

5. After assigning the substation capacities to the leaf nodes, the maximum possible power flows through all the edges of the network are obtained. Line losses are ignored at this stage, as the conductor properties are not yet assigned to the distribution lines. A demand factor $k_d = 0.6$, is assumed for all the secondary substations.

This step of the algorithm is explained using a simple example in Figure 5.7. The secondary subsection capacities are marked inside the boxes next to them. It is assumed that the direction of power flows in the network is from the source node to the leaf nodes. Then, the network is divided into a number of feeders equal to the degree of the source node. According to Figure 5.7, there are two such feeders: feeder 1 and feeder 2. The values obtained for the maximum possible power flows through all the edges in feeder 1 are shown in Table 5.8.

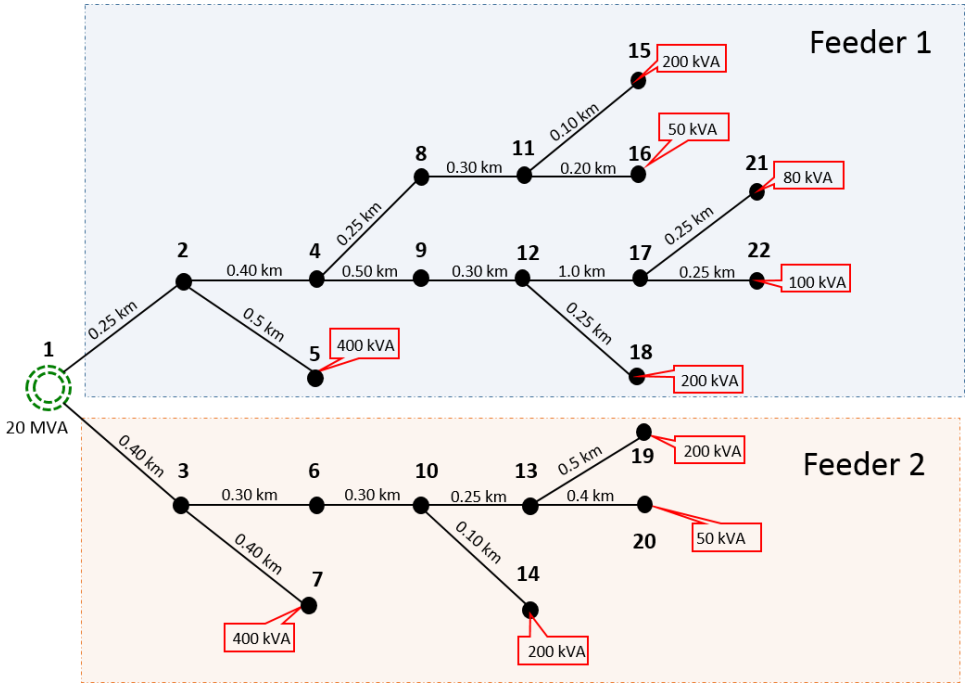


Figure 5.7: Assignment of electrical parameters to a distribution network topology.

For example, the maximum possible power flow through the edge $e_{x \to y}$ is the summation of all the installed capacities, $C_{node_{id}}$ of the leaf nodes in that specific feeder, with a $node_{id} \geq y$, multiplied by the demand factor k_d , as shown in Equation (5.10).

$$S_{x \to y} = k_d \times \sum_{node_{id} \geq y} C_{node_{id}} \tag{5.10}$$

Table 5.7: Maximum possible power flows through the edges of feeder 1 of the network.

Edge $e_{x \to y}$	$\sum C_{node_{id}}$ for $node_{id} \geq y$ (kVA)	Max possible power flow $S_{x \to y}$ (kVA)
$e_{1 \to 2}$	(400+200+50+80+100+200) = 1030	618
$e_{2 \to 4}$	(200+50+80+100+200) = 630	378
$e_{2 \to 5}$	400	240
$e_{4 \to 8}$	(200+50) = 250	150
$e_{4 \to 9}$	(200+80+100) = 380	228
$e_{8 \to 11}$	(200+50) = 250	150
$e_{9 \to 12}$	(200+80+100) = 380	228
$e_{11 \to 15}$	200	120
$e_{11 \to 16}$	50	30
$e_{12 \to 17}$	(80+100) = 180	108
$e_{12 \to 18}$	200	120
$e_{17 \to 21}$	80	48
$e_{17 \to 22}$	100	60

After assigning the substation capacities to the leaf nodes and computing the maximum branch flows in the network, now it is the time to choose the conductor cross sections for the distribution lines. Assuming that, all the distribution lines in the network have a uniform cross section is not a realistic approach.

In the traditional approach for planning and design of distribution networks, the selection of conductor cross sections comes after a series of other steps: forecasting of electricity demand, locating and sizing substations, designing the layout of the power distribution network [66].

In reality, the optimal conductor selection in distribution networks is an optimisation problem with the objectives to minimize the overall cost of annual energy losses and maximize the saving in the capital cost of conducting material, while satisfying the maximum current carrying capacity of the conductors and maintaining the acceptable voltage levels of network [71], [80], [81]. In addition, the effect of load growth over the years is also taken in to account [71].

However, the algorithm presented in this study does not go into the optimisation of costs and the energy losses when selecting the conductor cross sections for the distribution lines. The idea is to assign a similar selection of conductor cross sections to that of the real networks, while ensuring that the thermal limits and voltage drop constraints are not violated in the networks.

From the investigation in Chapter 4, it was observed that the trunk lines of the distribution networks are assigned with comparatively larger cross sections than the lateral branches. And also, it was observed that only 3-4 different types of conductor cross sections are used in one distribution feeder. These observations are used to implement the steps 6 -8 in this algorithm for the assignment of conductor cross sections to the distribution lines.

6. As mentioned before, identifying the trunk lines and laterals in the network is important in the process of selecting the right conductor sizes for the distribution lines.

   According to this step, each feeder has only one trunk line, the other edges are laterals. To identify the edges that belong to the trunk line, the following assumptions are made.

   (i) A trunk line is starting from the source node and ending at a leaf node in the corresponding feeder.

   (ii) However, a trunk line does not always appear from the source node to the leaf node which has the longest path length in kilo metres.

   (iii) The length of the path and the amount of power carried through the path must contribute in making the decision to identify the trunk line from the candidate paths.

   (iv) The calculation of power losses that occur in a candidate path for the trunk line involves both the lengths of the edges and actual power flows through the

edges. Therefore, the candidate path with the highest amount of power losses is chosen as the trunk line of the feeder.

The assumption (iv) is explained in detail below.

The active power loss, $PLoss_{e_{x\to y}}$ occurs in an edge $e_{x\to y}$ can be calculated using Equation (5.11). $P_{x\to y}$, $Q_{x\to y}$ and $S_{x\to y}$ represents the active, reactive and apparent power flows through the edge $e_{x\to y}$ respectively. Resistance and reactance of the edge are represented by R and X. Resistance, R of the edge can be substituted with $r \times l_{x,y}$ where r is the per-km resistance of the conductor and $l_{x,y}$ is the length of the edge. V_x and V_y are the node voltages.

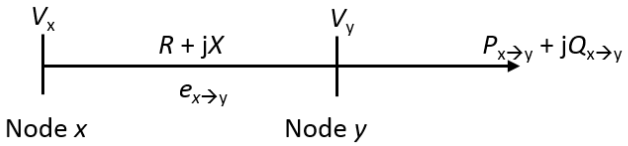


Figure 5.8: Power flow through an edge/branch of the network.

$$\begin{aligned} PLoss_{e_{x\to y}} &= \frac{P_{x\to y}^2 + Q_{x\to y}^2}{V_y^2} \times R \\ &= \frac{S_{x\to y}^2}{V_y^2} \times R \\ &= \frac{S_{x\to y}^2}{V_y^2} \times l_{x,y} \times r \end{aligned} \tag{5.11}$$

$$\begin{aligned} PLoss_{path\_p} &= \frac{r}{V_y^2} \times \sum_{\substack{All\ edges\ e_{x\to y} \\ in\ path\_p}} S_{x\to y}^2 \times l_{x,y} \\ &\cong Constant \times \sum_{\substack{All\ edges\ e_{x\to y} \\ in\ path\_p}} S_{x\to y}^2 \times l_{x,y} \end{aligned} \tag{5.12}$$

The total active power loss that occurs in the path *p* can be obtained from Equation (5.12). The maximum possible power flows in the edges obtained from the previous step (Step 5) are substituted to the $S_{x\rightarrow y}$ terms in the equation. The per-km resistance, *r* and the node voltages, V_y can be assumed to be constant values (In a 10kV distribution network the small deviations of the voltages at the nodes can be ignored for this calculation). Therefore, the total power loss occurs in the path *p* becomes proportional to the term, $\sum_{\substack{All\ edges\ e_{x\rightarrow y} \\ in\ path_p}} S_{x\rightarrow y}^2 \times l_{x,y}$. Which is called the 'loss index' of the path in this study.

Table 5.8 summarises the Step 6 of the algorithm with relevant to the feeder 1 of the example network shown in Figure 5.7. According to Table 5.8, path p_6 has the highest loss index. Hence, the edges (branches) that belong to the path p_6 become the edges that belong to the trunk line of the feeder.

Table 5.8: Selection of the candidate path for the trunk line of a distribution feeder.

Leaf node *id*	Path to the leaf node from the source node	Total length of the path	Loss index $\sum S^2_{x\rightarrow y} \times l_{x,y}$
5	p_1= (1 → 2 → 5)	0.75 km	175481
15	p_2= (1 → 2 → 4 → 8 → 11 → 15)	1.3 km	166450
16	p_3= (1 → 2 → 4 → 8 → 11 → 16)	1.4 km	165100
18	p_4= (1 → 2 → 4 → 9 → 12 → 18)	1.7 km	197820
21	p_5= (1 → 2 → 4 → 9 → 12 → 17 → 21)	2.7 km	206460
22	p_6= (1 → 2 → 4 → 9 → 12 → 17 → 22)	2.7 km	206790

7. After assigning the edges to the trunk lines, the remaining edges become laterals of the network. According to the capacity of the main supply transformer as given by the user, a set of conductor cross sections are chosen from the planning and design guidelines (Table 5.3). According to network planning and design guidelines it is recommended to use only three to four different cross sections for one feeder [75]. Therefore, the

selected set of conductor cross sections are divided into groups of three cross sections according to the ascending order of the cross sections as shown in Figure 5.9.

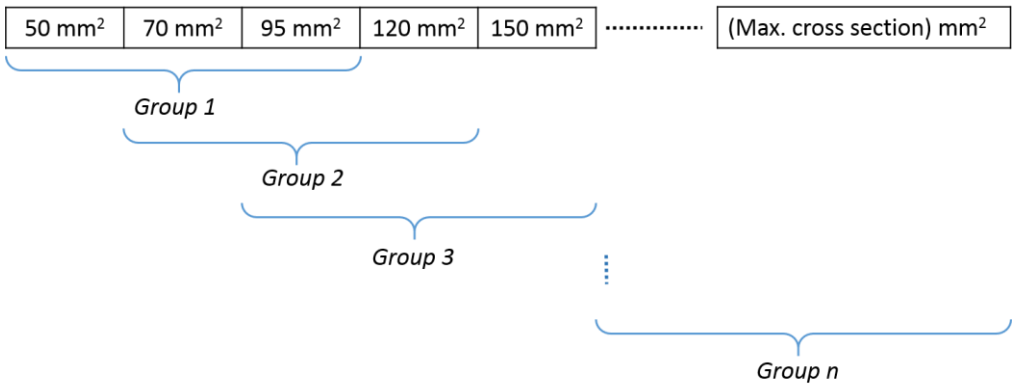


Figure 5.9: Grouping of conductor cross sections in ascending order.

Assignment of the cross sections to the branches in all the feeders begins with the group of smallest three cross sections. First, the largest cross section in the selected set is assigned to the edges in the trunk lines. The remaining two cross sections are assigned to the edges that belong to lateral branches. Then, the maximum power flows through the edges are tested against the thermal constraints of the selected conductors. If there are any violations of the thermal constraints the next set of three conductor cross sections are assigned to the edges of the feeders where the violations of thermal constraints are observed. This step is repeated until the thermal constraints are satisfied in all the edges of the network.

8. In this step of the algorithm, a power flow simulation is conducted to obtain the maximum voltage drop recorded in the network. The maximum recorded voltage drop is compared with the maximum allowable voltage drop of the network (V_{d_max}). If the constraint for the maximum allowable voltage drop is violated, the algorithm goes back to the previous step to choose the next set of conductor cross sections (i.e. cross sections larger than in the previous assignment). These steps will be repeated until the voltage drop constraints are satisfied or the combinations of all the recommended sets

of cross sections are finished. If there is no suitable set of conductor cross sections that can satisfy the voltage drop constraint, the network is discarded. Otherwise the network is stored in the database of statistically-similar networks.

9. One saved network has the following output files: Adjacency matrix, $\boldsymbol{A}$ Electrical distance matrix, $\boldsymbol{ED}$ secondary substations matrix, $\boldsymbol{SS}$ and branch matrix, $\boldsymbol{B}$. The construction of $\boldsymbol{SS}$ and $\boldsymbol{ED}$ is shown in Equations (4. 2) and (4.12) respectively. The structure of $\boldsymbol{B}$ is as shown in Equation (5.13). It contains $From_{node_{id}}$ and $To_{node_{id}}$ of the branches, the length l*(km)* of the branches, per-km resistance r (Ohm), per-km reactance x (Ohm) *and* also the thermal capacity $C_{thermal}$ (kVA) of the conductors used in the branches. M is the total number of branches of the network. These output files are used to conduct power flow studies of the networks.

$$\boldsymbol{B} = [From_{node_{id}}, \quad To_{node_{id}}, \quad l, \quad r, \quad x, \quad C_{thermal}]_{M\times 6} \tag{5.13}$$

5.5 Validation of the SSNG tool

The steps of the validation procedure of the SSNG tool is shown in Figure 5.1. This section presents the validation of the SSNG tool by choosing one network from the *testing* data set. This selected network is referred to as 'test network' in the text below.

First, an ensemble of statistically-similar distribution networks to the selected test network is generated by the SSNG. Then the topological and electrical properties of the test network are compared with the topological and electrical properties of the ensemble of statistically-similar networks generated by the SSNG. The results of the comparison are discussed.

5.5.1 Properties of the test network sample

The basic inputs taken from the test network to provide 'direct user inputs' to the SSNG is summarised in Table 5.9. The population density of the selected network is 620 people per square km, and hence it belongs to the sub-urban category of networks.

Table 5.9: Basic user inputs from the test network to the SSNG.

Property	Value
Network type	Sub-urban
Total number of nodes	190
Capacity of the main supply transformer	20 MVA
Number of outgoing feeders from the main supply transformer	5

The topological and electrical properties of the test network is calculated by following the methodologies described in Chapters 3 and 4. The results are shown in Table 5.10.

Table 5.10: Topological and electrical properties of the test network.

Topological properties		**Electrical properties**	
Property	Value	Property	Value
Total network length	62.53 km	Total installed capacity of secondary substations (kVA)	16875 kVA
Average edge length	0.331 km	Average capacity of a secondary substation	219.2 kVA
Average path length	6.49 km	Average electrical distance-resistance	0.274 Ohms
Maximum feeder length	8.87	Average electrical distance-reactance	0.125 Ohms
Average node degree	1.989	Total active power loss	340 kW
Branching rate	0.349	Total reactive power loss	194 kVar
Number of leaf nodes	77	Total power loss/total supplied power ratio (%)	3.72 %
		Minimum recorded voltage	0.982 p.u. @ bus 170
		System load balancing index	0.1554

Figure 5.10(a) shows the spatial distribution of the actual test network. Figure 5.10(b) shows the connections of the nodes in the actual network as a simplified tree diagram. The edge lengths in Figure 5.10 (b) are not drawn according to an actual scale. The trunk lines of the network are marked using red colour in Figure 5.10(b).

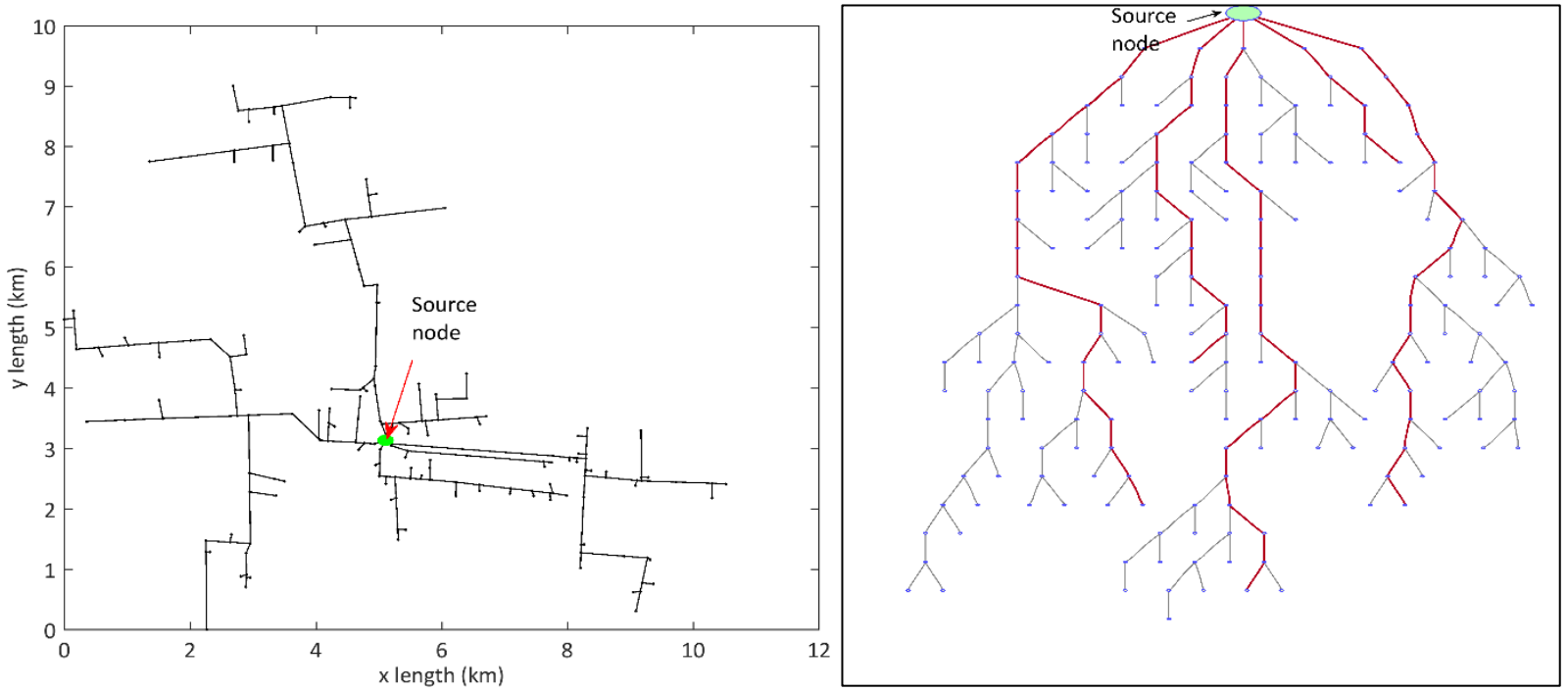


Figure 5.10: (a) Spatial distribution of the actual test network (b) Connections between the nodes of the test network as a simplified tree diagram.

5.5.2 Generation of statistically-similar networks

Since, the proposed networks generator is a statistical tool, it can generate large numbers of statistically-similar networks. In this section, the basic inputs in Table 5.9 were used to generate 100 networks which are statistically-similar to the chosen network from the *testing* dataset. Studies with different numbers of statistically-similar networks can be conducted see the effect of the number of networks on the statistics of the results. However, this part of the study is not covered in this thesis.

SSNG model parameters that come from the real-world networks investigation for topology generation are shown in Table 5.11. SSNG model parameters to assign electrical properties to the network topologies are as follows. (i) A demand factor of k_d= 0.6, is assumed for secondary substations to estimate the maximum possible branch flows. (ii) The minimum allowable voltage drop in the networks is considered as 10% of from main supply point voltage. This limit is about 6% in the process of planning and design of real-world networks [66]. But from the network investigation in the Chapter 4, with a similar set assumption made to carry out load flow simulations a voltage drops of more than 6% is observed in the real-world networks. Hence, this voltage drop constraint is relaxed up to 10% for this study.

Table 5.11: SSNG model parameters for topology generation.

Level of the network,i	**Parameters for edge length distributions**	**Parameters for degree distributions**	**Parameters for distribution of the nodes**
	α_i	$[p_{i1}, p_{i2}, p_{i3}]$	q_i
$i = 1$	0.324	[0.32 0.35 0.33]	0.475
$i = 2$	0.264	[0.39 0.28 0.33]	0.375
$i = 3$	0.236	[0.45 0.28 0.27]	0.150

Figure 5.11 shows the simplified tree diagrams of three networks generated by the SSNG. Again, the edge lengths are not drawn according to the actual scale. Trunk lines are marked in red colour.

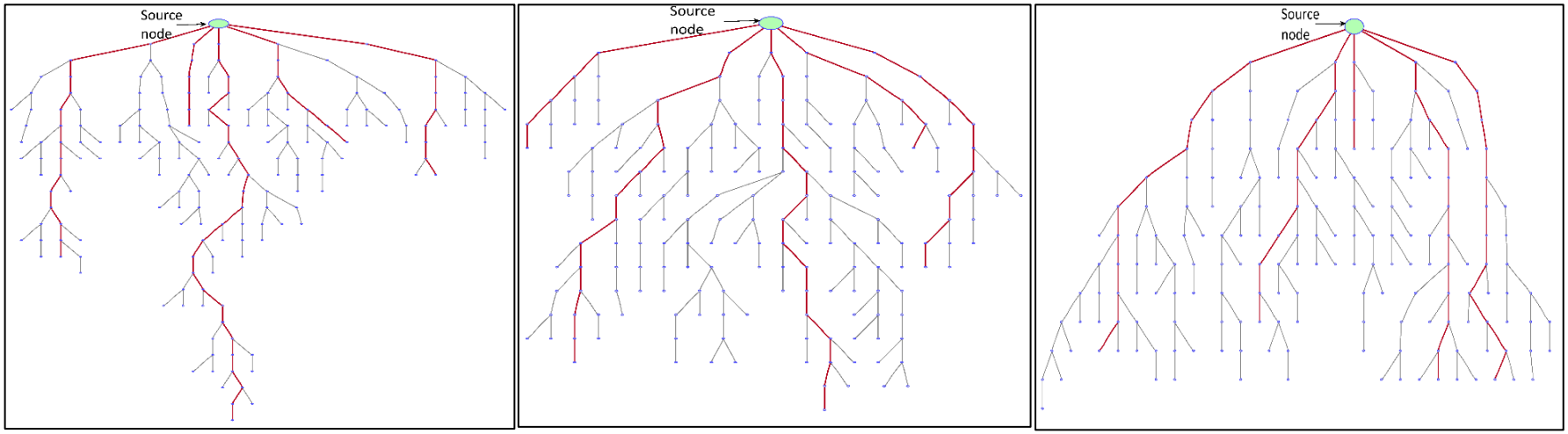


Figure 5.11: Layouts (as simplified tree diagrams) of three statistically-similar networks generated by the SSNG.

5.5.3 Comparison of the topological and electrical properties of the test network with the networks generated by the SSNG.

The results obtained for the topological and electrical properties of 10 networks out of the 100 networks generated by the SSNG are listed in Table 5.12. The mean and standard deviations of the topological and electrical properties of 10 networks are also calculated and compared with the test network properties in the last three rows of Table 5.12.

To get a better understanding about the representativeness of the networks generated by the SSNG to the test network, results of all the 100 networks generated by the SSNG are considered. The box-whisker plot representations of the topological and electrical properties of the 100 networks are shown in Figure 5.12. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol [82].

All the networks generated by the SSNG have same number of total nodes and same number of outgoing feeders from the source node similar to the test network.

From the results in Table 5.12 and Figure 5.12 it can be observed that most of the topological and electrical properties (excluding the electrical performance related electrical properties) of the generated networks have a standard deviation less than or equal to 12%. This observation proves that the networks generated by the SSNG are statistically-similar to each other in terms of those properties with a maximum range of variation around ±12%.

Table 5.12: Topological and electrical properties of 10 statistically similar networks.

Network number	Topological properties							Electrical properties								
	Total network length (km)	Average edge length (km)	Average path length (km)	Maximum feeder length (km)	Average node degree	Branching rate	Number of leaf nodes	Total installed capacity of secondary substations (kVA)	Average capacity of a secondary substation (kVA)	Average electrical distance- resistance (Ohms)	Average electrical distance- reactance (Ohms)	Total active power loss (kW)	Total reactive power loss (kVar)	Minimum recorded voltage (pu)	System load balancing index	Total power loss/total supplied power ratio (%)
1	65.26	0.345	4.38	5.72	1.989	0.342	69	13915	201.7	0.299	0.131	188	95	1.034	0.0917	1.62
2	60.34	0.319	5.11	8.1	1.989	0.311	63	13020	206.7	0.191	0.117	208	179	1	0.1087	3.4
3	66.28	0.351	4.96	6.72	1.989	0.353	71	15865	223.5	0.215	0.129	124	103	1.025	0.0858	1.67
4	58.08	0.307	3.97	5.8	1.989	0.347	70	15290	218.4	0.184	0.112	132	91	1.032	0.0922	1.72
5	60.22	0.319	4.61	5.45	1.989	0.347	70	12150	173.6	0.286	0.121	155	72	1.012	0.0935	2.28
6	58.32	0.309	5.03	6.21	1.989	0.347	70	14655	209.4	0.133	0.11	224	137	1.023	0.0724	2.07
7	64.03	0.339	5.57	5.99	1.989	0.337	68	15690	230.7	0.308	0.129	243	160	1.02	0.0918	2.57
8	63.34	0.335	4.07	5.34	1.989	0.342	69	13010	188.6	0.305	0.127	146	61	1.024	0.1123	1.98
9	61.43	0.325	5.09	6.66	1.989	0.337	68	12245	180.1	0.286	0.123	232	117	0.987	0.1024	3.42
10	66.84	0.354	5.53	6.34	1.989	0.337	68	13860	203.8	0.153	0.126	131	129	1.019	0.0735	2.16
Mean	62.41	0.330	4.83	6.23	1.989	0.340	69	13970	203.7	0.236	0.122	178	114	1.017	0.0924	2.29
Std. dev, ±	5%	5%	12%	12%	0%	3%	3%	9%	9%	28%	6%	25%	32%	1.4%	13%	27%
Test network	62.53	0.331	6.49	8.87	1.989	0.349	77	16875	219.2	0.274	0.125	340	194	0.982	0.1554	3.72

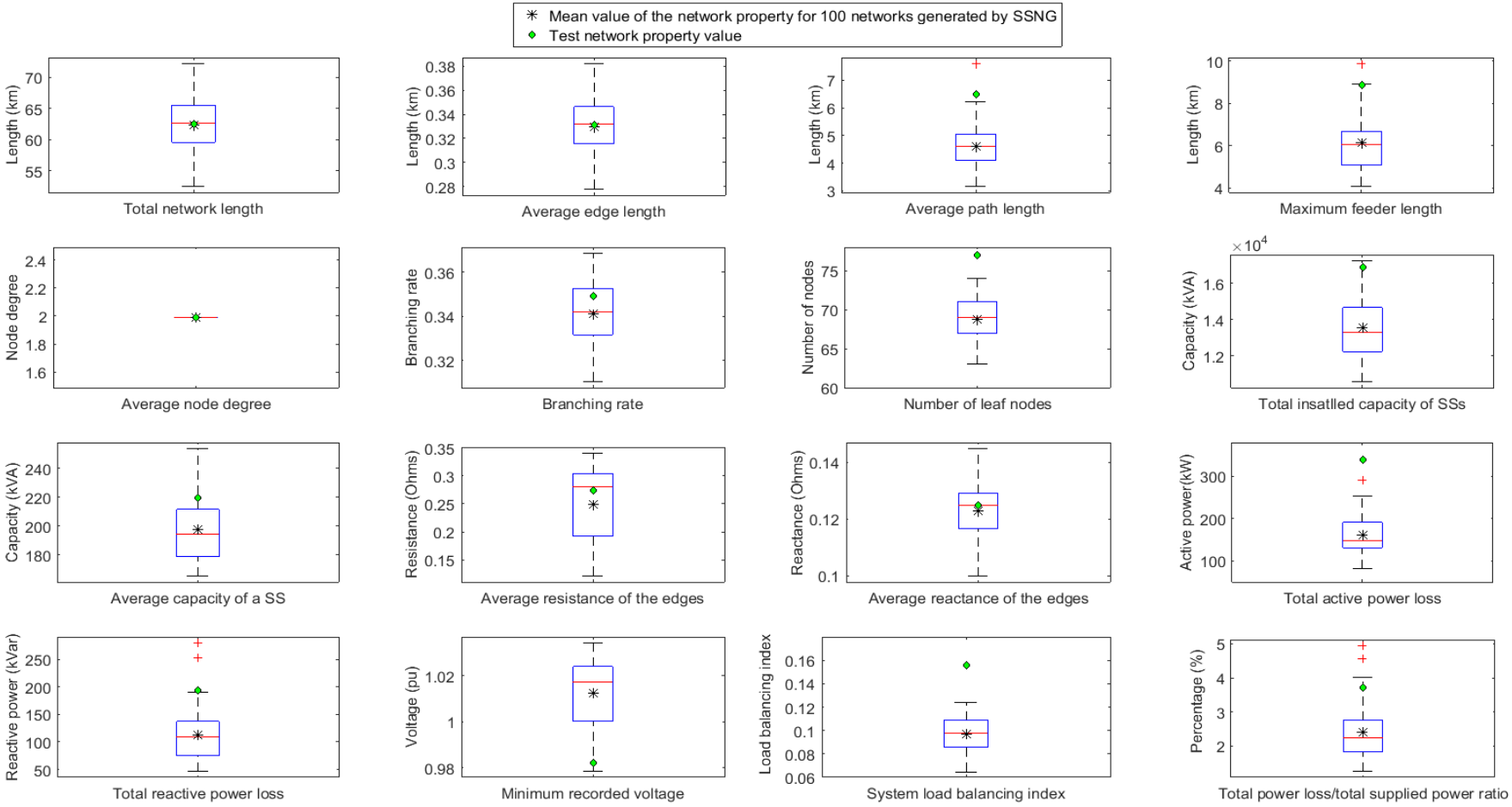


Figure 5.12: Box-whisker representation of the topological and electrical properties of 100 statistically-similar networks.

From the results in Table 5.12 and Figure 5.12 it can be observed that, the mean values of the network samples generated by the SSNG, for the topological properties such as, total network length, average edge length, average node degree and branching rate are very similar to the corresponding property value of the test network.

The difference between the mean value for the number of leaf nodes (equals to the number of secondary substations) in the generated networks and the test network is around 10%. Whereas the mean values of the average path length and maximum feeder length deviate in a range of 25-30% from the corresponding property values of the test network.

The mean value for the average capacity of a secondary substation in the generated networks are about 10% lower than that of the test network. The test network property values for the average resistance and reactance of the distribution line segments (edges) are in the range of 25^{th} and 75^{th} percentiles of the statistical distribution of the properties of generated networks.

However the electrical performance related properties of the generated networks such as power losses, load balancing index and minimum recorded voltage are considerably varied from the test network properties.

In summary, the networks generated by the SSNG are very closely representative to the test network, in terms of the total network length, average edge length and branching properties. The number of secondary substations and average capacity of the secondary substations are also nearly similar to the test network. Deviations from the electrical performances related network properties such as, power losses, voltage drops and load balancing indexes are observed with compared to the test network.

5.5.4 Discussion-Fine tuning the algorithms for networks generation

Deviations observed in some of the electrical and topological properties of generated networks and the test network may be due to the effect of the following reasons.

(i) Impact of the population density of the selected network

The selected test network has a population density of 620 people per square km. This much larger than most of the sub urban networks used in the investigation of the real-world network properties. The higher the population density the networks tend to have more branching to supply to consumers leading to more leaf nodes in the network. And also the installed capacities of the secondary substations tend to be larger to supply more customers. However, the SSNG model parameters used in this study to generate the networks are derived using a majority of sub-urban networks with a population density less than the selected test network.

(ii) Mismatches in the selection of conductor cross sections of the test network and the networks generated by the SSNG.

From the results in Table 5.12 and Figure 5.12, it is observable that the networks generated by the SSNG has lower branch impedances than the test network. This may be due to idealistic assumptions made to obtain the maximum branch flows in the networks generated by the SSNG (Step 5 of the algorithm for electrical property assignment).

The SSNG can be fine-tuned to get better representative networks to the test network. This can be done by changing the SSNG model parameters. The SSNG model parameters for topology generation (Table 5.12) can be varied to change the topological structures of the networks generated by the SSNG. For example, by changing the parameter (α) related to the edge length

distributions of the networks, different total network lengths can be achieved. Likewise, by changing the degree distributions of the networks related parameters (p), different branching patterns of the networks can be achieved.

All these small changes of the SSNG model parameters for topology generation can be justified, as they are derived using the properties of the real-world networks and those properties of the real networks can be slightly varied from one network to the other even within the same network type according to the population distribution.

Similarly, the SSNG model parameters for assigning the electrical parameters can also be fine-tuned. The assumptions for the demand factor can be changed achieve the branch flows which will eventually give a similar selection of conductor cross sections to the generated networks, similar to that of the test network.

However, fine-tuning the SSNG is not implemented as an in-built function to the model. It is a heuristic approach at this stage of the development of SSNG.

5.6 Summary

A statistically-similar network generator (SSNG) for MV electricity distribution networks was developed. The topological and electrical properties in real-world networks together with the guidelines for distribution network planning and design are used in the development process of the SSNG.

Statistically similar networks are generated in a hierarchical way. Algorithms are developed; first, to generate realistic topologies of distribution networks and then, to assign realistic electrical parameters to the network topologies.

The validation of the SSNG showed that the tool is capable of generating ensembles of statistically-similar networks which resemble the real-world networks in terms of a set of topological and electrical properties.

It has been identified that the population density parameter within the same type of networks has an impact of the detailed topological and electrical properties of the networks. The SSNG model parameters can be fine-tuned to improve the representativeness of the networks to the real-world networks.

Chapter 6: Case study- Impact assessment of Soft Open Points on distribution networks

The work presented in this Chapter is a collaboration with Miss. Qi Qi of Cardiff University, School of Engineering, UK.

6.1 Introduction

With a high penetration of intermittent DG and flexible demand presents in the distribution networks, there is a possibility that some parts of the network become heavily loaded at certain time of a day and lightly loaded at other times. This, in turn, leads to high power losses, increased peak currents and undesirable over-voltage issues in the distribution networks [83]. The use of power electronic devices provides alternative solutions to overcome these challenges [84], [85].

- **An introduction to the Soft Open Points (SOP)**

Soft Open Point (SOP) is a power electronic device installed in place of a normally open or normally closed point in a distribution network [86]. Normally Open Points (NOPs) are built, connecting adjacent feeders, to provide alternative routes of electricity supply in case of planned or unplanned power outages [87].

Instead of simply opening/closing NOPs, SOPs control load transfer and regulate network voltage profile by flexibly controlling active/reactive power flow between adjacent feeders. Immediate fault isolation between interconnected feeders as well as fast supply restoration is also enabled using these devices. Therefore, SOPs are able to improve distribution network

operation as well as facilitate a large penetration of low carbon technologies into the distribution network [86].

- **A literature review of research studies on SOPs**

A number of previous studies have investigated the benefits of SOPs for distribution network operation [83], [86], [88], [89], [90], [91]. This previous research on SOP has mainly focused on one or few of the following aspects: (i) development of control strategies for SOP; (ii) minimization of network energy losses considering the influence and increase of DG; (iii) analysis and quantification of benefits of SOP considering different objectives separately [92]. To the best of the author's knowledge, all the previous studies on SOPs have been validated or tested against one or few case study networks, such as IEEE 33 busbar network.

Therefore, these previous research studies leave the question mark of how applicable and relevant the benefits of SOPs quantified by that specific study for the other real-world networks.

- **Goals and novelty of the current research.**

The overall benefits that can bring to the distribution network by the SOPs can be varied from one network to the other. Therefore, it is important to test the SOP's capability of bringing advantages on many different distribution networks in order to provide broad and robust conclusions about the benefits of SOPs on distribution networks.

As an initial step forward to address this research gap, a methodology for the investigation of the SOP's capability of bringing benefits to the distribution networks on many realistic distribution networks is proposed in this chapter.

An integrated tool for this purpose is developed by combining the SSNG which was presented in the Chapter 5 of this thesis and the optimization framework that was proposed in the reference [92], by Qi Qi of Cardiff University, UK.

The techniques of obtaining statistically meaningful results from a large number of networks, about the SOP performances are demonstrated using a set of statistically similar networks. SOP operation in the distribution networks at different penetration levels of the DGs are considered.

6.2 Statistical assessment tool for performance evaluation of SOP on distribution networks.

This section describes the implementation of the integrated tool for SOP performance evaluation on distribution networks. The integrated tool is referred to as the 'Statistical assessment tool for the performance evaluation of SOP on distribution networks'.

The mathematical model of the SOP used in a distribution network in this study is explained. The details of the optimization framework that was proposed to improve the operation of a distribution network with distributed generation and a SOP, are presented.

6.2.1 An overview of the integrated tool for SOP performance evaluation

Figure 6.1 shows the simulation set up of the integrated tool for SOP performance evaluation. A similar arrangement for the impact assessment of solar PV on the LV distribution networks is presented in [93].

The steps of the simulation study are as follows. First, a realistic distribution network is generated by the SSNG considering the user inputs of the network to be generated. A DG penetration level is selected. Then DG sites are randomly distributed in the network and the

capacities for the DGs are assigned according to the DG penetration level. The presence of only one SOP per network is considered in this study. Then, the SOP performance evaluation model (an optimisation algorithm) is executed to obtain the outcomes for the optimal capacity (kVA) of SOP, optimal location of SOP and annual savings gained by using the SOP. The results are recorded. Then another case is run with the same DG locations for the same distribution network, but with a different DG penetration level. This is continued until the number of DG penetration levels required for the SOP study has been completed. Then, another network is generated by the SSNG and all the above steps are repeated until enough cases have been generated to provide a meaningful spread of results.

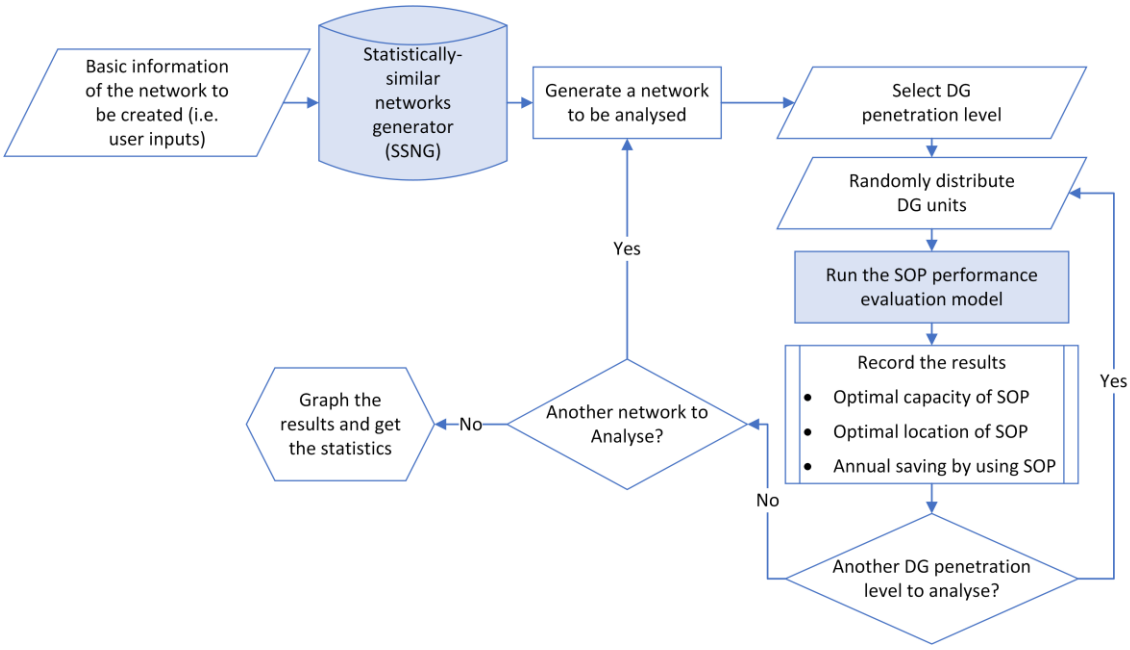


Figure 6.1: Simulation setup of the statistical assessment tool for SOP performance evaluation [93].

Different sets of user inputs regarding the networks that are required for the SOP study can be given to the SSNG. However, in this chapter only one set of statistically-similar networks are used for SOP performance evaluation.

6.2.2 DG consideration in the distribution network models

This research is focused on evaluating the performances of SOPs in the distribution networks in the presence of DGs. The details of the considerations made for DG allocation in the distribution network models is explained in this section. The locations for the DG sites are chosen randomly.

Although there is no unified definition of DG penetration in the literature, this study follows the commonly used one, which is the same as that in [86], [94]: DG penetration is the percentage of total capacity of DG units over the maximum loading capacity of the network. For example, if a distribution network is operating at its peak load with a DG penetration of 25%, then 75% of the power will be coming from the transmission system.

In order to consider the intermittent characteristics of renewable generation outputs, a method based on the Wasserstein distance was used to derive representative scenarios of DG outputs from their probability density functions [95], [96]. Firstly, the probability density functions of DGs were generated from historical data. Then the continuous distributions of DG outputs were converted to discrete ones using the Wasserstein distance method, from where representative operation scenarios of DGs over a certain period were generated. By using these representative scenarios of DG operation, the computational burden of the optimization procedure (i.e. optimisation to improve the operation of a distribution network with a SOP) can be reduced.

Assuming $f(x)$ is the continuous probability density function of variable x, and is to be converted into N_S discrete distributions, i.e. N_S representative scenarios. Each representative scenario sce ($sce = 1,2, \dots, N_s$) and its corresponding probability p_{sce} can be obtained as follows:

$$\int_{-\infty}^{sce} f(x)^{1/2}dx = \frac{sce-1}{2N_S}\int_{-\infty}^{+\infty} f(x)^{1/2}dx \tag{6.1}$$

$$p_{sce} = \int_{\frac{(sce-1)+sce}{2}}^{\frac{sce+(sce+1)}{2}} f(x)dx \tag{6.2}$$

Specifically, the scenarios $sce = 0$ and $sce = N_s + 1$ refer to the lower and upper limits of the variable x. In this study x, refers to the renewable generation in the form of wind and PV. Since Weibull distribution [97] and Beta distribution [98] are extensively used to describe the probability density functions of wind speed and light intensity in the literature, they were adopted for the generation of representative scenarios of wind and PV outputs over a year, based on the Wasserstein distance method.

The wind speed and light intensity curves over a year as shown in Figure 6.2, and the consequent probability density functions derived from these curves were obtained in [96]. The probability density functions were then used with the Wasserstein distance method to derive five representative scenarios of wind outputs (P_{wind}), five representative scenarios of PV outputs (P_{PV}), as well as their corresponding probabilities. From these scenarios, 25 combined scenarios of annual DG operation, and their probabilities can be obtained, which are shown in Table 6.1.

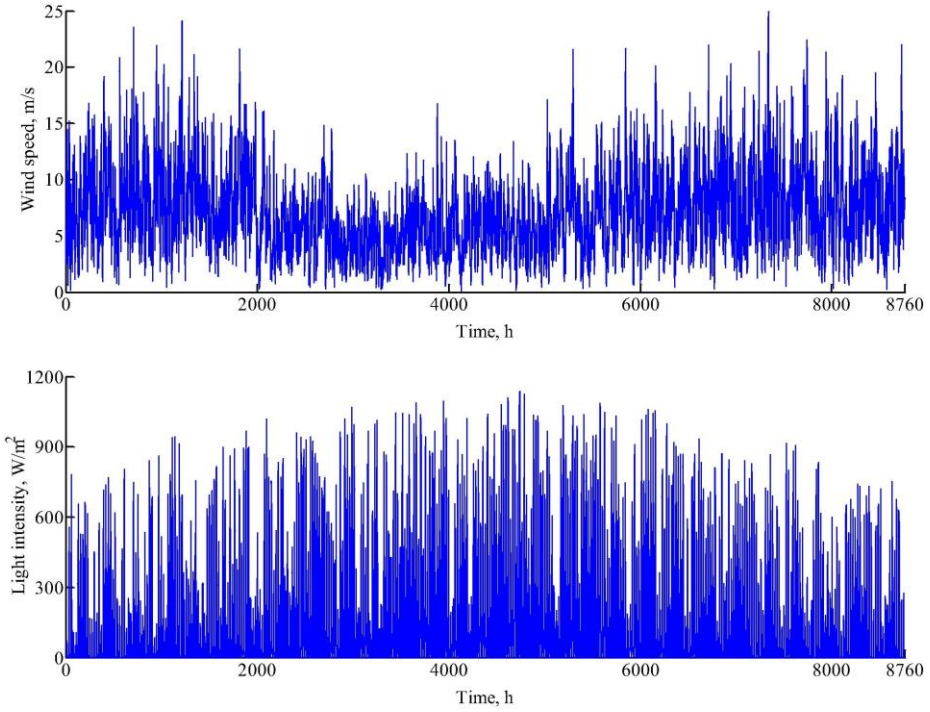


Figure 6.2: Wind speed and light intensity curves over a year.

Table 6.1: Representative scenarios and corresponding probabilities of DG outputs.

P_{wind}		P_{PV}				
		sce 1	*sce* 2	*sce* 3	*sce* 4	*sce* 5
		0.09	0.29	0.49	0.72	0.91
***sce* 1**	0	0.018	0.0174	0.0173	0.0174	0.018
***sce* 2**	0.08	0.0916	0.0885	0.0882	0.0885	0.0915
***sce* 3**	0.32	0.0468	0.0452	0.0451	0.0452	0.0468
***sce* 4**	0.71	0.0268	0.0259	0.0258	0.0259	0.0268
***sce* 5**	1	0.021	0.0203	0.0203	0.0203	0.021

6.2.3 Mathematical model of SOP in distribution networks

A SOP can be implemented with different converter topologies. A back-to-back Voltage Source Converter (VSC) based SOP was considered in this study. A schematic diagram of a distribution network installed with an SOP is given in Figure 6.3 [92].

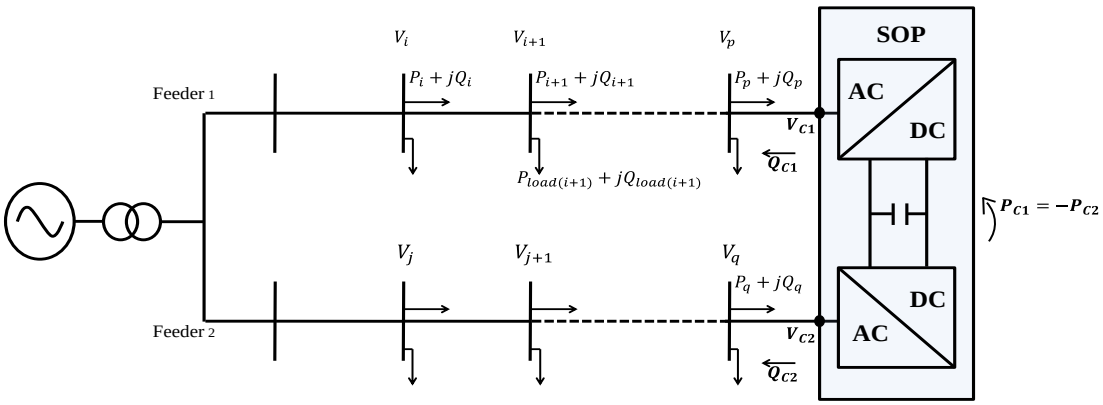


Figure 6.3: A distribution network installed with an SOP [92].

Back-to-back VSCs can operate in four quadrants of the P-Q capability curve [91]. The reactive power at both terminals of an SOP can be assigned independently to each other as required, which makes the device capable to provide flexible reactive power compensation to the network. In addition, the active power flow through an SOP can be controlled rapidly and accurately.

To fully evaluate the potential effects of SOP on steady-state network operations, a mathematical power injection model of SOP is used [91]. In this model, active and reactive power injections at SOP terminals are integrated into the load flow algorithm without considering the detailed design of converter controllers. The backward forward sweep method was used for load flow calculations. Taking Feeder 1 in Figure 6.3 as an example, the load flow is calculated by the following recursive equations [99]:

$$P_{i+1} = P_i - P_{loss(i,i+1)} - P_{load(i+1)} = P_i - \frac{r_i}{|V_i|^2} \cdot \left(P_i^2 + Q_i^2\right) - P_{load(i+1)} \tag{6.3}$$

$$Q_{i+1} = Q_i - Q_{loss(i,i+1)} - Q_{load(i+1)} = Q_i - \frac{x_i}{|V_i|^2} \cdot \left(P_i^2 + Q_i^2\right) - Q_{load(i+1)} \tag{6.4}$$

$$|V_{i+1}|^2 = |V_i|^2 + \frac{r_i^2 + x_i^2}{|V_i|^2} \cdot \left(P_i^2 + Q_i^2\right) - 2 \cdot (r_i P_i + x_i Q_i) \quad i \in \{1,2,\dots,N_{bus}\} \tag{6.5}$$

where, P_i and Q_i are the active and reactive power flowing from bus i to bus $i+1$. $P_{load(i)}$ and $Q_{load(i)}$ are the active and reactive power demand at bus i. $P_{loss(i,i+1)}$ and $Q_{loss(i,i+1)}$ are the power losses within the branch connecting buses i and $i+1$, and r_i and x_i are the resistance and reactance of that branch. V_i is the voltage at bus i. N_{bus} is the total number of buses in a network.

The operational boundaries of a back-back VSC-based SOP are:

$$P_{C1} = P_p - P_{loss(p,C1)} \tag{6.6}$$

$$P_{C2} = P_q - P_{loss(q,C2)} \tag{6.7}$$

where, $P_{C1}$and P_{C2} are the active power flows of each VSC. $P_{loss(p,C1)}$ is the power loss between bus p and VSC 1, and $P_{loss(q,C2)}$ is the power loss between bus q and VSC 2.

Although the operation efficiency of back-back VSCs is sufficiently high, they inevitably produce losses when there is a large-scale transfer of active power. Therefore, a loss coefficient η is considered in the model. The constraint of active power exchange between the two converters is described below:

$$P_{C1} + P_{C2} + \eta P_{C1} + \eta P_{C2} = 0 \tag{6.8}$$

The reactive power outputs of the two converters are independent of each other because of the DC isolation and should satisfy their own capacity constraints:

$$\sqrt{P_{C1}^2 + Q_{C1}^2} \leq S_{SOP} \tag{6.9}$$

$$\sqrt{P_{C2}^2 + Q_{C2}^2} \leq S_{SOP} \tag{6.10}$$

where, Q_{C1}, and Q_{C2} are the reactive power output of each converter, and S_{SOP} is the capacity of an SOP.

The relationship between the capacity and location of an SOP is:

$$S_{SOP} = mS_{module}(1 - b_{i,j}) \tag{6.11}$$

where, the capacity of an SOP equals to the summation of the module capacity it consists of, where S_{module} is the minimum capacity of the basic power electronic module in an SOP. The binary variable $b_{i,j}$ indicates if the branch between buses i and j is equipped with an SOP. For instance, $b_{i,j} = 0$ means an SOP is installed that branch.

Generally, the AC side of an SOP can be controlled in either PV mode or PQ mode. In this study the latter was considered. By choosing the optimal capacity, location and set-points of an SOP, power flows within a network can be controlled actively. Therefore, specific operational objectives can be achieved.

6.2.4 SOP performance evaluation model

The SOP performance evaluation model is formulated as an optimisation problem, where the annual cost of a distribution network with a SOP and different penetration levels of DGs is subjected to minimise.

- **Problem formulation**

The annual cost of a distribution network was taken as the objective function of the optimisation problem and formulated as follows:

$$min\, F = C_{inv} + C_{ope} + C_{loss} \tag{6.12}$$

The annual cost consists of the following three parts:

(i) Investment cost of SOP:

$$C_{inv} = \frac{r_{inv}(1 + r_{inv})^{y}}{(1 + r_{inv})^{y} - 1} c_{SOP} S_{SOP} \tag{6.13}$$

where, r_{inv} is the discount factor for the investment cost of SOP, y is the device economical service time, c_{SOP} is the investment cost per unit capacity of an SOP.

(ii) Operational cost of SOP

$$C_{ope} = r_{ope} c_{SOP} S_{SOP} \tag{6.14}$$

where, r_{ope} is the discount factor for the operational cost of SOP.

(iii) Annual energy loss cost of a distribution network:

$$C_{loss} = 8760 \cdot c_{ele} \sum_{sce=1}^{N_S} \left(P_{loss,sce} + \eta P_{C1,sce} + \eta P_{C2,sce}\right) \cdot p_{sce} \tag{6.15}$$

where c_{ele} is the electricity price per kWh. N_S is the set of all scenarios considered over a year, and p_{sce} is the probability corresponding to scenario sce. $P_{loss,sce}$ is the power loss of the network per unit time at scenario sce. $P_{C1,sce}$ and $P_{C2,sce}$ are the active power exchanged through the converters of an SOP at scenario sce.

In addition to the operational constraints of SOP, the following limits were also considered:

(i) Bus Voltage Limits

$$V^{min} \leq |V_i| \leq V^{max} \quad i \in \{1,2,\dots,N_{bus}\} \tag{6.16}$$

(ii) Branch Capacity Limits

$$|I_k| \leq I_k^{max} \quad k \in \{1,2,\dots,N_{branch}\} \tag{6.17}$$

where, I_k^{max} is the maximum allowed current of branch k.

- **Optimization framework**

During the optimization process, the active and reactive power outputs of an SOP during each representative scenario, as well as the capacity and location of the SOP were taken as decision variables. Decision variables are a set of quantities that need to be determined in order to solve the problem. Their values are searched and identified in the optimization model in order to obtain the optimal value of the objective function.

An integrated optimization method proposed in the reference [92], was used to determine the optimal allocation of SOP by considering the values obtained for the decision variables. This method integrates both global and local search techniques, where Particle Swarm Optimization

(PSO) [100] is adopted to explore the solution space globally. A local search technique, namely the Taxi-cab method [101], is used for solution space exploitation, which refines the quality of solutions searched by PSO in each iteration, overcoming the drawback of PSO in local optima trapping. Therefore, the search capability of the integrated optimization method has enhanced search capability compared to the conventional PSO algorithm.

After identifying the optimal solution, the annual cost savings obtained by the SOP is calculated as follows,

$$\begin{aligned}\text{Annual Cost Saving}(\%) &= \frac{\mathrm{F}^{\mathrm{ref}} - \mathrm{F}^{\mathrm{SOP}}}{\mathrm{F}^{\mathrm{ref}}} \times 100\% \\ &= \frac{\mathrm{C}_{\mathrm{loss}}^{\mathrm{ref}} - \left(\mathrm{C}_{\mathrm{inv}}^{\mathrm{SOP}} + \mathrm{C}_{\mathrm{ope}}^{\mathrm{SOP}} + \mathrm{C}_{\mathrm{loss}}^{\mathrm{SOP}}\right)}{\mathrm{C}_{\mathrm{loss}}^{\mathrm{ref}}} \cdot 100\%\end{aligned} \tag{6.18}$$

$\mathrm{F}^{\mathrm{ref}}$ is the annual cost of the distribution network in the reference case, where no SOP is installed. $\mathrm{F}^{\mathrm{SOP}}$ is the annual cost of the distribution network with one SOP installed and $\mathrm{C}_{\mathrm{loss}}^{\mathrm{ref}}$ stands for the annual energy loss cost of the distribution network in the reference case, where no SOP is installed.

6.3 Case study

The statistical assessment tool proposed for SOP performance evaluation in Section 6.2, is used to construct a case study. A set of statistically-similar distribution networks are tested for their performance with a SOP and different DG penetration levels.

6.3.1 Description of the test cases

- **Input parameters given to the SSNG**

The set of input parameters that are given to the SSNG in this case study, are shown in Table 6.2

Table 6.2: The set of input parameters given to the SSNG.

Parameter	Value
Network Type	Sub-urban
Voltage level of the network	10kV
Number of nodes	250
Number of outgoing feeders from the main transformer	7
Capacity of the main transformer	20 MVA
Number of statistically-similar networks required for the study	30

- **Outputs generated from the SSNG**

Each statistically-similar distribution network model generated by the SSNG have several output files. The output file 'Branch matrix', ***B*** includes the connectivity of nodes (buses) in the network, branch impedances and current rating of the conductors used in the branches. Secondary substations matrix, ***SS*** gives the installed capacities of the secondary substations connected to the nodes (buses) in the network. Buses and branches that belong to the trunk lines of each separate feeders of a distribution network models are also given as an output file from the SSNG.

- **Assumptions made for modelling the consumer load in the distribution networks**

The optimisation model for the SOP performance evaluation, takes into account the consumer load information of the network. The optimisation model is run for a period of one year. However, the SSNG is only capable of providing the information of the distribution of installed capacities of secondary substations in the network models. Time variant load profiles are not generated by the SSNG.

Therefore, in order to model the consumer loads in the distribution networks models, the following assumptions are made. The nodes where the secondary substations are connected,

are considered as the load nodes of the distribution network models. A fraction equals to 0.6 of the installed capacities of secondary substations is considered as the instantaneous load connected to the corresponding node at any time of the year. It is also assumed that the all the loads are operating at a power factor of 0.85.

- **DG consideration in the distribution networks**

DG sites are allocated randomly in each network. All DGs were assumed to operate at unity power factor. Since the installed capacity of DGs is related to their penetration levels, in this study, DG penetrations from 0 to 100% with an increment of 20% are considered.

- **SOP information**

The consideration of the range of candidate SOP capacities was from 500 kVA to 2500 kVA. End nodes of the trunk lines of each main feeder of the distribution networks are considered as candidate points for SOP installation. Out of all such candidate locations for SOPs, the optimisation algorithm selects the best location for a SOP for each network to satisfy the objective of the study.

- **Parameters used in the optimisation**

Parameters related to the investment and operational costs of SOP used in the optimisation problem, are listed in Table 6.3.

Table 6.3: Parameters selected for the case studies.

Parameters	Value
loss coefficient: η [102]	0.02
module capacity: S_{module} (kVA)	100
economical service time: y (year)	20
discount factor: r_{inv}, r_{ope}	0.08, 0.01

investment cost per unit capacity: c_{SOP} (£/kVA) [103], [104]	230
electricity price: c_{ele} (£/kWh)	0.12

6.3.2 Results and discussion

The box-whisker plot representations of the topological and electrical properties of the 30 networks generated by the SSNG are shown in Figure 6.4.

From the box-whiskers plots in Figure 6.4, it can be observed that the networks generated by the SSNG share very close topological and electrical properties making them statistically-similar to each other. However, these networks have different network layouts and any of that is a possible realisation of a real distribution network.

The difference in the network layouts with the random integration of the DGs can lead to different performances of the statistically-similar networks.

Under each DG penetration level, for each network, the annual network costs, network performances (E.g. maximum and minimum bus voltages recorded over a year, maximum branch loading recorded over a year) and the maximum allowable DG penetration of the network before any limits being breached are obtained for the cases without SOP and with SOP operation in the networks.

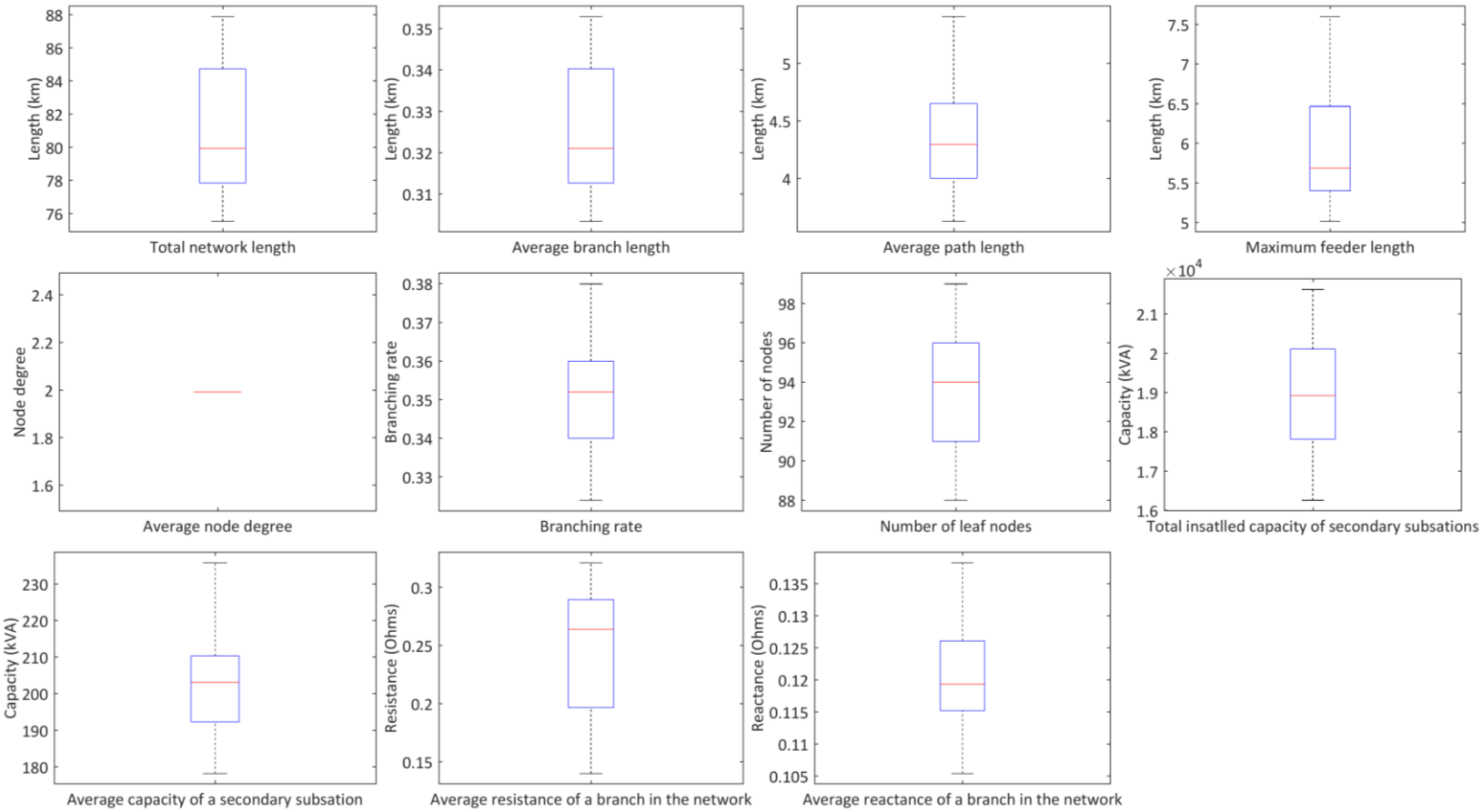


Figure 6.4: Topological and electrical properties of 30 networks generated by the SSNG.

- **Annual costs of distribution networks with and without SOP**

The annual costs of the 30 statistically-similar distribution networks under different DG penetration levels are compared for the cases of with and without SOP in the networks. Figure 6.5 shows the comparison of the annual costs using a box plot representation. Networks with high DG penetrations are more vulnerable to encounter voltage and thermal issues. Therefore, only the results of the networks without constraints violations are used for the representation.

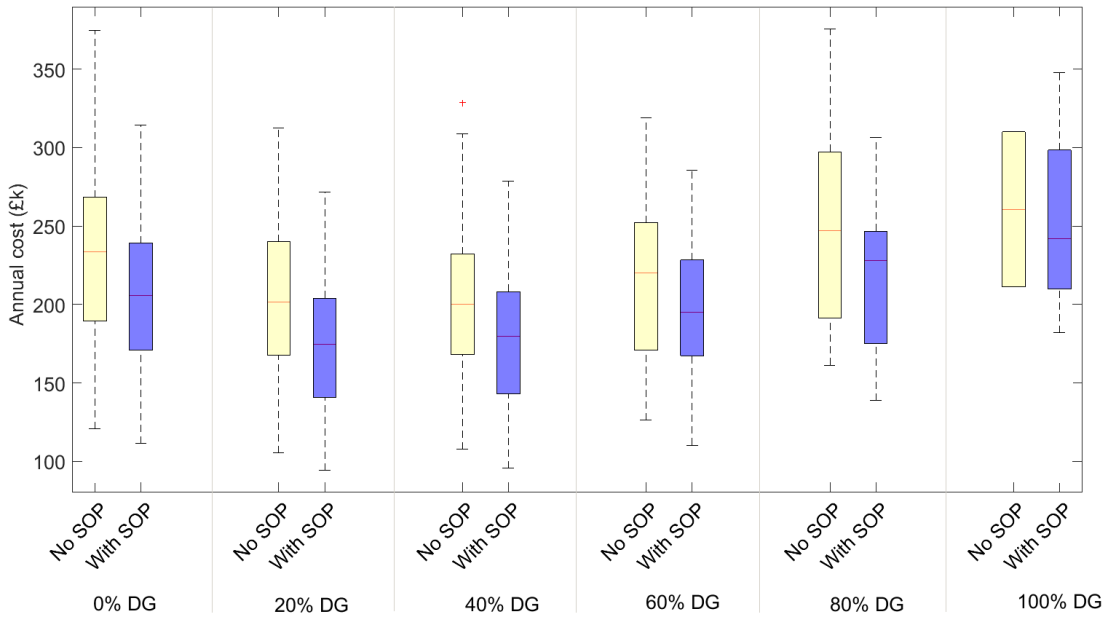


Figure 6.5: Annual costs of 30 statistically-similar distribution networks, with and without SOP, for different DG penetration levels.

From Figure 6.5 it can be observed that for the both cases (i.e. with and without the SOP), with an increasing penetration levels of DGs initially the network annual costs has reduced. However, later when the DG penetration levels continues to increase the annual costs have started to increase. This variation of the annual costs of networks with different DG penetration levels can be explained using the energy losses occur in the networks.

For the cases without SOP, the annual costs of the networks are purely defined by the energy losses occur in the networks. For the cases with SOP, in addition to the cost of energy losses the investment and operational costs of the SOPs are also involved in calculating the annual costs of the networks. With reverse power flows being introduced by the DGs, for lower DG penetration levels the overloading of some of the feeders can be reduced. This can lead to reduced energy losses and hence reduced network costs can be observed. However, with higher DG penetration levels, DG can also introduce overloading for some of the feeders in the networks. This leads to higher energy losses and hence higher annual costs are observed for high DG penetration levels.

It can be clearly observed from Figure 6.5, that the annual costs of networks with SOP is always less than those of the cases without using SOP, for all the DG penetration levels.

- **Annual savings brought by using a SOP in distribution networks**

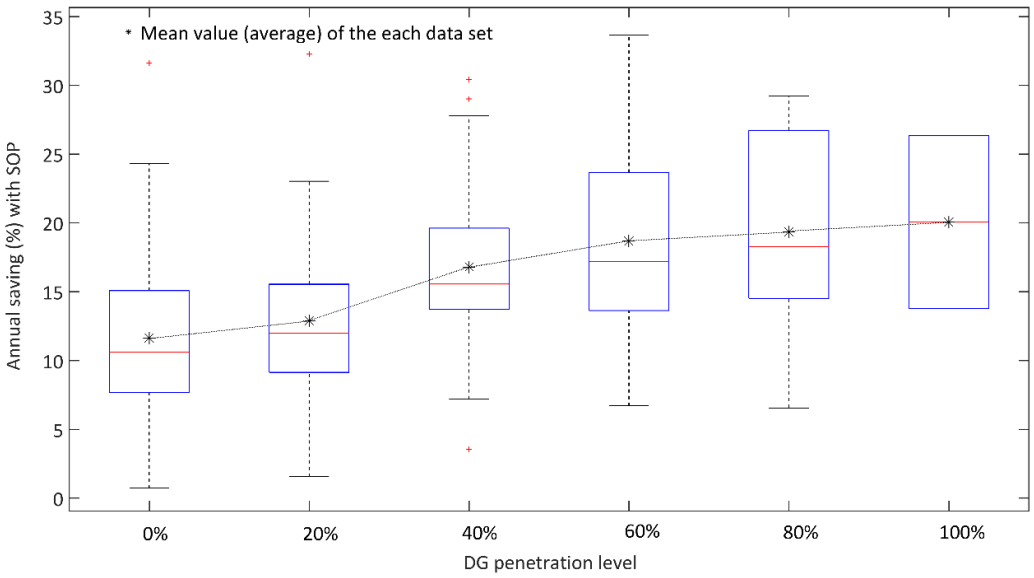


Figure 6.6: Annual costs savings brought by SOP for 30 statistically-similar networks under different DG penetration levels. The outliers are plotted using the '+' (red) cross marks.

The annual savings brought by the SOPs for the distribution networks with different DG penetration levels are calculated using Equation (6.18). The statistics of the 30 statistically-

similar networks for the annual savings are represented by box plots shown in the Figure 6.6. Mean value of the annual cost savings for each DG penetration is shown by the '*' mark in each box plot. According to Figure 6.6, with the increment of the DG penetration the annual cost savings brought by the SOP has increased slowly.

The annual savings calculated in terms of a percentage (%), allow more general conclusions to be made about the cost benefits of SOPs. For the type of network that has been considered in this study, with different network layouts and random DG locations the annual cost savings brought by SOP is in the range of 10-20%, for a wide range of DG penetration from 0-100%.

- **Comparison of network performances with and without SOP**

The Minimum and maximum recorded voltages (p.u.) over a period of one year in the statistically-similar distribution networks, with and without SOP for different DG penetration levels are compared in Figure 6.7. Results of the networks without constraints violations are presented here. Jittering has been introduced to the visualisation to prevent over-plotting in statistical graphics (when visualising a large number of data points, jittering is used to add a little random noise to the data in order to see the data points in a cloud more clearly) [105].

For very high DG penetration levels the number of data points observed in Figure 6.7 have been reduced due to the voltage and thermal constraint violations of the networks. According to Figure 6.7, with increasing DG penetration levels the maximum recorded voltages in the networks have clearly increased for the cases without SOP. However, the maximum recorded voltages in the networks have been reduced considerably in all DG penetration levels by using SOPs. It can be also observed that the minimum recorded voltages in the networks have been improved by using SOPs in all DG penetration levels. These observations prove that SOP is

capable of providing voltage regulation in the selected type of distribution networks for a wide range of DG penetration levels.

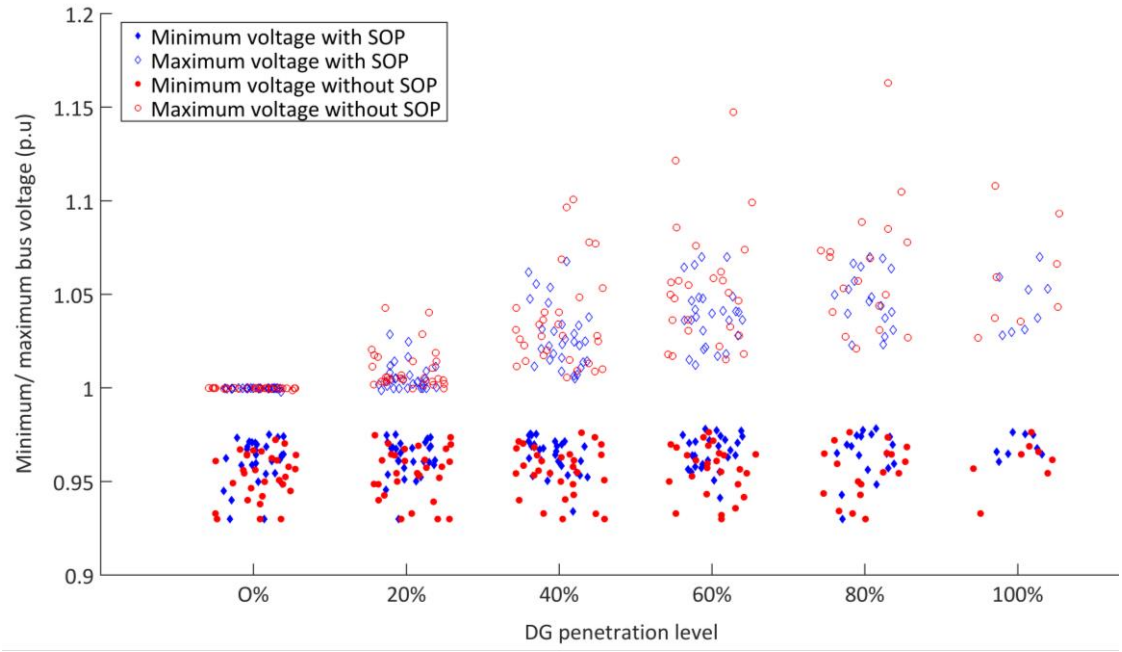
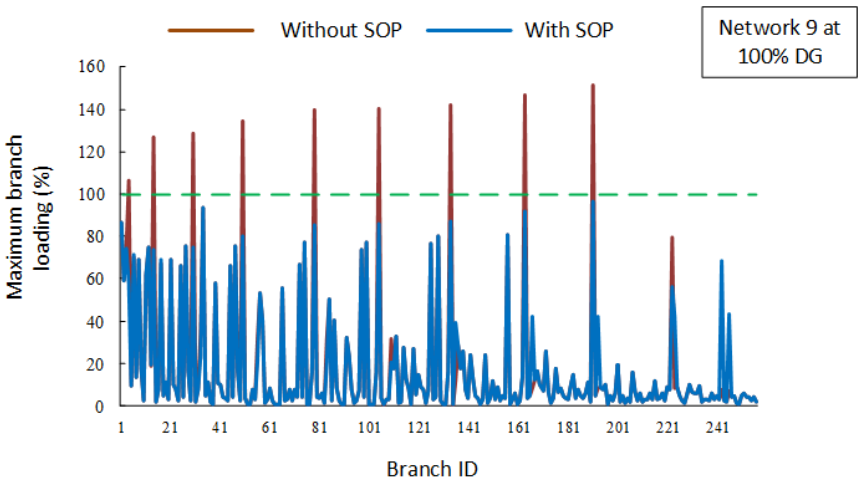


Figure 6.7: The maximum and minimum bus voltages recorded over a year under different DG penetration levels.

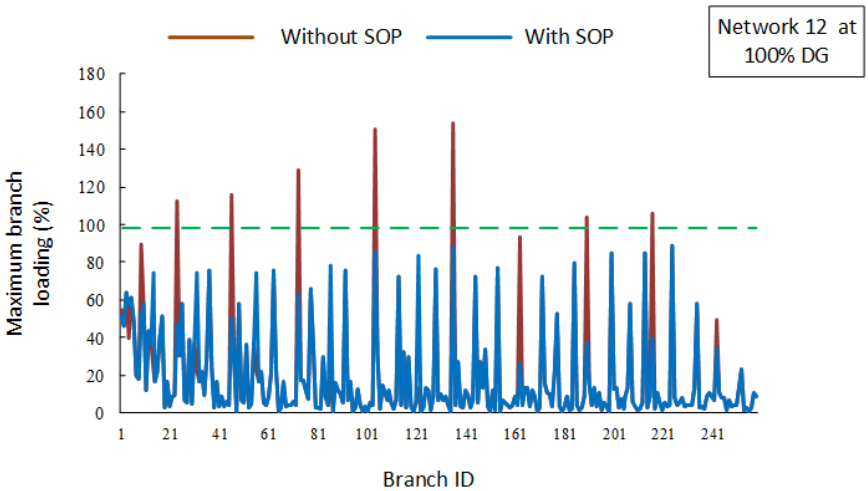
Next, the maximum branch loadings observed in the distribution networks for the cases of with and without SOPs are compared. Maximum branch loading (%) for each branch of the network is obtained by dividing the maximum power flow through a specific branch at a given time by the maximum thermal capacity of that branch.

The maximum branch loadings recorded in each network over one year period of time is considered. At first, two networks, which have great improvement in the allowable DG penetration level without violating the voltage and thermal constraints are used for the discussion. The two networks used for the results in Figure 6.8, allow a maximum of 100% DG penetration level with a SOP whereas the maximum allowable penetration for those two networks without a SOP is only 60%.

In Figures 6.8(a) and 6.8(b) the red lines represent the maximum branch loadings without SOP and blue lines represent the maximum branch loadings with SOP for a DG integration level of 100%. Green dashed line shows the margin of the branch overloading (i.e. 100% loading of a branch). As seen from the figures SOP is capable of reducing the maximum loading of the branches thus allowing more DG integration to the networks.



(a)



(b)

Figure 6.8: Maximum branch loadings recorded over a year for the networks achieving great improvements in DG integration by using SOPs.

Figures 6.9(a) and 6.9(b) show the maximum branch loading recorded over a year, for two networks achieving no improvements in DG integration when using SOPs. The maximum allowable DG penetration level with and without SOP for these two networks is 80%.

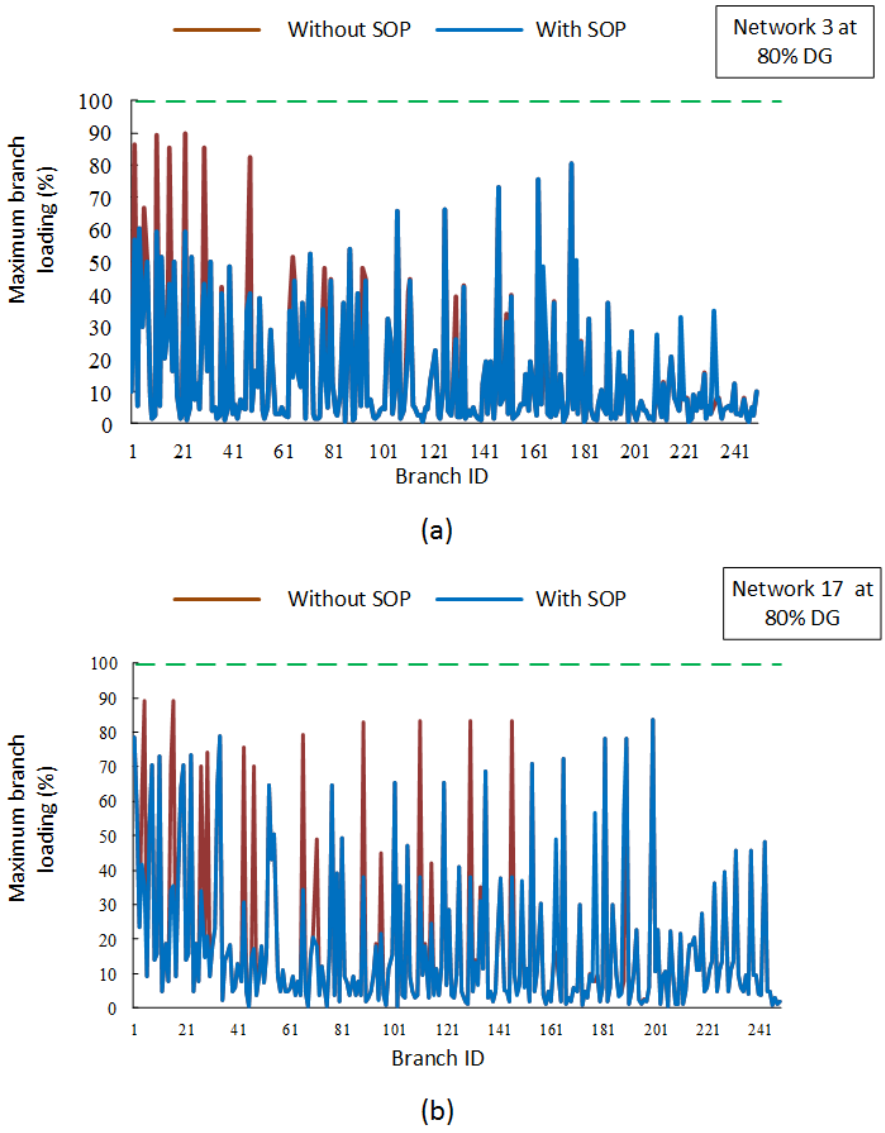


Figure 6.9: Maximum branch loadings recorded over a year for the networks achieving no improvement in DG integration by using SOPs.

According to Figures 6.9(a) and 6.9(b) although there is no improvement of the maximum allowable DG integration level by using a SOP, the branch loading of both networks has been reduced by using SOPs.

Therefore, it is evident from the results that the SOPs are capable of reducing the branch loadings of the distribution networks. However, an improvement of the allowable DG integration level cannot be always expected by using a SOP in the type of distribution networks selected for this study. The network constraints that limit the DG integration is closely defined by the factors such as, network layout, consumer load distribution and the locations of the DGs.

- **Maximum allowable DG integration with and without SOP**

The maximum allowable DG penetration levels of the 30 statistically-similar distribution networks, for the cases of with and without SOPs, are shown using a box plot representation in Figure 6.10. Average values of the maximum allowable DG penetration level with and without SOP are shown using the '*' marks within the corresponding box plot. Results show that SOPs can facilitate further DG integration to the selected type of distribution networks. As a general conclusion, for the selected type of distribution networks an average of 20% increment of the DG integration level can be expected by using SOPs.

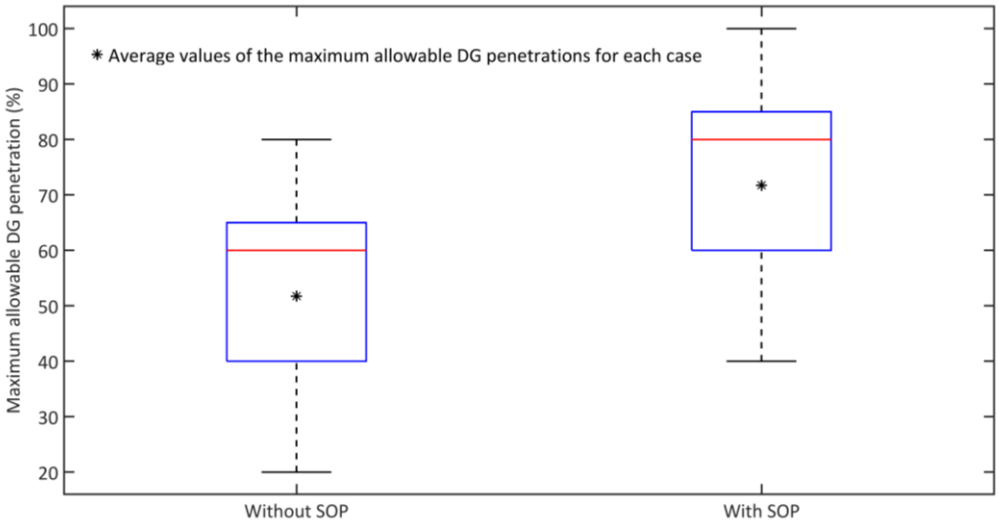


Figure 6.10: Maximum allowable DG penetration levels of the 30 statistically-similar networks with and without SOPs.

- **Required optimal capacity of the SOP.**

The optimal capacities required for the SOPs for each network at different DG penetration levels are shown in Figure 6.11. The number of dots shown for each DG penetration level in Figure 6.11 is not always equal to the number of networks tested. This is because, some networks violate the voltage and thermal constraints at certain integration levels of DGs even with SOPs. The average capacity of the SOP at each DG penetration level is marked using a red star within the corresponding data cloud.

From the results it can be observed that, for a small DG penetration level of 20%, the required capacity of the SOP is less than that of the DG integration level of 0%. This is due to the reduced energy losses brought by small levels of DG integrations. However, for larger DG penetration levels (i.e. DG penetration level >20%), the required capacities of SOPs have increased with the DG penetration level.

From the results it can be concluded that, a SOP with the capacity under 2.5 MVA is always sufficient for the type of distribution networks used for this study, for the DG integration levels of 0-100%.

Figure 6.11: Required optimal capacities of SOPs at different DG penetration levels of the 30 statistically-similar networks.

6.4 Summary

A network assessment tool for the evaluation of the benefits of SOPs on distribution networks is proposed in this chapter. Using the proposed tool, the feasibility of performing a benefit analysis of SOPs on a set of statistically-similar networks is presented.

Operation of SOPs in the distribution networks at different DG penetration levels is considered. An optimisation problem which minimises the annual cost of the distribution network is solved to obtain the optimum capacities of the SOPs at each DG penetration level of the distribution network. The network performances, in terms of minimum and maximum recorded voltages and maximum branch loadings are compared for the cases of without SOP and with SOP in the networks. Maximum allowable DG penetration levels of the networks with and without SOPs are also obtained.

Results of the 30 statistically-similar networks showed that, for the type of distribution networks used in this study,

(i) the annual cost of the distribution network with SOP is always less than cost of the distribution networks without SOP. The annual cost savings that can obtain by using a SOP in the distribution networks is around 10—20%;

(ii) with high penetration levels of DGs, over voltage and thermal overloading issues in the networks are increased. SOPs are able to provide voltage and power flow controls to reduce or even to eliminate these issues while facilitating further DG integration to the networks;

(iii) by using a SOP, the maximum allowable DG penetration level in the networks is increased approximately by a 20% with compared to the cases without SOPs; and

(iv) the required capacity of a SOP tends to increase for the higher DG penetration levels. A SOP with a maximum capacity of 2.5 MVA is sufficient for the type of distribution network under study, for a wide range of DG integration from 0 – 100%.

Chapter 7: Conclusions and future work

This chapter provides an overview of the research undertaken, the conclusions and recommendations for the future work.

7.1 Conclusions

The overall aim of this thesis is to develop a statistical assessment tool for electricity distribution networks which can facilitate a large number of simulation studies on many realistic distribution networks in order to come up with robust and generalised conclusions on impact studies of LCTs. As parts of this thesis the following objectives were set out.

(i) To review existing literature on analysing statistical properties of power networks and the available methods and tools for modelling the power networks.

(ii) To investigate the topological properties of real-world electricity distribution networks.

(iii) To investigate the electrical properties of real-world electricity distribution networks.

(iv) To develop a statistically similar networks generator (SSNG) for electricity distribution networks.

(v) To demonstrate the application of the SSNG tool to analyse the impacts of soft open points on EDNs with variable levels of DG penetration.

7.1.1 Review of impact studies of LCTs, statistical studies on power networks and methods and tools for power network modelling.

The importance of robust decision making in the face of significant uncertainties in the electrical power networks have been discussed. The different levels of decision stakes (operational, strategic and policy level) in the power systems are discussed.

A review of the literature on the impact assessment studies of LCTs of the distribution networks was conducted.

- It has been identified that most of the previous studies on the impact analysis of various LCTs were conducted on one or few case study networks. Results of those studies have a limited applicability to the other networks making it difficult to arrive to generalised conclusions.
- It was recognised that one of the main reasons for using a case study or synthetic networks in those research was limited accessibility to real-world network data. Therefore, the requirement of a network modelling tools which can generate large numbers of realistic distribution networks was identified.

A review on the previous research studies on analysing the statistical properties of the real-world power networks and the available network modelling methods and tools for power networks was undertaken.

- Outcomes of the review showed that, most of the previous research on analysing the topological and electrical properties of real world power networks was carried out in the HV level of the power network and the studies carried out on the MV and LV networks are very limited.

- Well-known complex network models have been used to model the HV power grid. The structure of HV grids is different from that of MV and LV grids. The HV transmission and sub-transmission is usually a meshed system, but distribution networks (MV and LV) are mainly with radial structures. Therefore, the research findings in the HV network analysis cannot be directly used in MV and LV networks.
- A few fractal based network models have been developed to model the distribution networks. However, some of these studies were supported by a limited set of real network data or a limited set of properties of the real networks were analysed.
- The lack of network modelling tools which can generate random-realistic representations for electricity distribution networks has been identified.
- The requirement of a comprehensive statistical investigation of the topological and electrical structures of the electricity distribution networks in order to develop network modelling tools for electricity distribution networks have been identified.

7.1.2 Investigation of topological properties of MV electricity distribution networks

An investigation of the topological properties of the MV real-world networks was conducted using network data collected from China, covering urban and sub-urban areas. Only radial structures of the networks were considered in the study.

The motivations behind the topological investigation presented in this chapter were to find out

(i) the key topological properties that characterize the realistic nature of different types of electricity distribution networks, and

(ii) a possible way to efficiently generate ensembles of random but realistic network topologies similar to real electricity distribution networks.

The techniques from complex networks analysis and graph theory were employed in the investigation.

A novel approach to obtain depth-dependent topological properties has also been developed.

Results of the investigation of topological properties in real-world networks showed that,

- node degree and edge length related graph properties are fundamental in characterizing the topological structures of radial type sub-urban and urban electricity distribution networks; and
- depth dependent properties were able to better capture the topological features of electricity networks at different depth levels of the networks. Results from the depth dependent analysis showed that urban and sub-urban types of electricity distribution networks have different graph related properties at different depth levels of the networks.

7.1.3 Investigation of electrical properties of MV electricity distribution networks

Electrical properties of real-world electricity distribution networks at the MV level were studied. The same set of Chinese network data that was used in the topological investigation was used for the study. The motivation of this study was to identify and quantify the key electrical properties which will later be useful in network modelling.

A limited set of available data regarding the installed capacities of distribution substations and the conductor cross sections of the distribution lines were used for the study.

A novel approach to obtain depth dependent electrical properties has also been developed.

Results from the real-world network investigation showed that

- the substations capacities and the conductor cross sections are able to characterise the electrical features of sub-urban and urban networks; and
- Kernel density PDFs which describe the distribution of secondary substation capacities along the feeder lengths of the two types of networks have clear differences and reflects the realistic consumer distributions of the two types of networks.

7.1.4 Development of the Statistically-Similar Networks Generator

A statistically-similar network generator (SSNG) for MV electricity distribution networks was developed. The results from the investigations of topological and electrical properties of real-world networks together with the guidelines for distribution network planning and design were used in the development process of the SSNG. Therefore SSNG is a data driven tool.

The ability to generate statistically-similar many networks to a replicate one distribution network is one of the key features of the SSNG.

The network generation was done in a hierarchical way. First, a realistic topology of a distribution network was generated and then, realistic electrical parameters are assigned to the network topology. This process was repeated to generate a large number of realistic distribution networks.

The validation of the SSNG was done by comparing the topological and electrical properties of a real-world test network with an ensemble of statistically-similar networks which were generated to replicate the selected real-world test network.

- The validation of the SSNG showed that, the SSNG tool is capable of generating ensembles of statistically-similar networks which resemble the real-world networks in terms of a set of topological and electrical properties.
- It has been identified that the population density parameter within the same type of networks has an impact of the detailed topological and electrical properties of the networks.
- The ability to fine-tune the SSNG model parameters to improve the representativeness of the networks to the real-world networks was discussed.

7.1.5 Impact assessment of SOPs on distribution networks to increase the DG penetration level.

A network assessment tool was developed by combining the SSNG tool and the SOP performance evaluation model which was developed by Qi Qi of Cardiff University, UK.

The feasibility of performing a benefit analysis of SOPs on a set of statistically-similar networks was presented by using the developed network assessment tool. The statistics of the results of the generated set of networks was used to provide general conclusions about the type of the networks used for the study.

Operation of SOPs in the distribution networks at different DG penetration levels was considered. Maximum allowable DG penetration levels of the networks with and without SOPs were also obtained.

Results of the set statistically-similar networks showed that, for the type of distribution networks used in this study,

- the annual cost of the distribution network with SOP is always less than cost of the distribution networks without SOP. The annual cost savings that can obtain by using a SOP in the distribution networks is around 10—20%;
- with high penetration levels of DGs, over voltage and thermal overloading issues in the networks are increased. SOPs are able to provide voltage and power flow controls to reduce or even to eliminate these issues while facilitating further DG integration to the networks;
- by using a SOP, the maximum allowable DG penetration level in the networks is increased approximately by a 20% with compared to the cases without SOPs, and
- the required capacity of a SOP tends to increase for the higher DG penetration levels. A SOP with a maximum capacity of 2.5 MVA is sufficient for the type of distribution network under study, for a wide range of DG integration from 0 – 100%.

7.2 Recommendations for future research

This research has provided the initial steps to a statistical platform where various network studies can be carried out on many statistically-similar realistic test networks.

The statistical assessment tool presented in this thesis is developed for 10kV Chinese distribution networks. SSNG tool can be further developed to enhance its performances.

The proposed future work includes,

- to conduct sensitivity analysis studies with different number of 'depth levels' in the investigations of the depth dependent topological and electrical properties of electricity distribution networks. The work presented in this thesis has not considered the impact of using different number of 'depth levels' in the investigations of network properties

and also in the development process of SSNG. Therefore, such sensitivity analysis will help to figure out whether there is an impact of the number of 'depth levels' on the accuracy and performance of the SSNG;

- to conduct sensitivity analysis with different numbers of statistically-similar networks, in order to identify the effect of the number of networks on the statistics of the results;
- to incorporate real/realistic load profiles of domestic, industrial and commercial consumers in to the load flow studies of statistically-similar networks. The case study of SOPs presented in this thesis, assumes a fraction equals to 0.6 of the installed capacities of secondary substations as the instantaneous load connected to the corresponding node at any time of the year. This is a limitation of the present study;
- to conduct impact assessment studies of various other LCTs such as solar PV and EV on the electricity distribution networks using the developed tool. This will allow further development of the statistical approach in conducting the research on many networks and in providing generalised conclusions of the results;
- to improve the SSNG tool with the real-world network data from different countries, different areas (rural, urban) and from different voltage levels including the LV networks. Also, the SSNG can be improved to generate both the radial and meshed type network structures. This will allow SSNG to generate network models for a wide range of real-world networks;
- to improve the visualisation of the networks in order to provide better representations of the real-world networks in terms of spatial structures. This will allow more constraint to be involved in the network generation such as actual area of the network, when producing realistic representations for the distribution networks. At the current stage of the development of SSNG it does not take into account the spatial constraints of the actual networks;

- to utilise the data available from GIS systems on the actual population distribution of the consumer areas in the generation of networks. From the investigations of topological and electrical properties of real world networks and it has been identified that the distribution of the population has a very close relation with the network properties, and
- to extend the idea of this research for the other energy carrier networks such as gas and heat distribution networks.

Appendix A: An introduction to Kernel Density Estimation

In statistics, kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable [76], [106].

A histogram is the simplest non-parametric density estimator. Kernel density estimates are closely related to histograms. Therefore, the construction of histograms and kernel density estimates can be compared to better understand the KDE. To construct a histogram, the interval covered by the data values is divided into equal sub-intervals, known as `bins'. Every time, a data value falls into a particular sub-interval, then a block, of size equal 1 by the bin width, is placed on top of it. When a histogram is constructed, two main points are considered: the size of the bins (the bin width) and the end points of the bins. Because of this bin count approach, the histogram produces a discrete probability density function. This might be unsuitable for certain applications, such as generating random numbers from a fitted distribution [106].

Unlike a histogram, which discretizes the data values into separate bins, the kernel distribution builds the probability density function by creating an individual probability density curve for each data value, then summing the smooth curves. This approach creates one smooth, continuous probability density function for the data set [106]. Figure A.1 shows a comparison of the histogram and kernel density estimate constructed using the same set of data shown in the Table A.1.

Table A. 1: A test data set.

Data Sample	Sample_1	Sample_2	Sample_3	Sample_4	Sample_5	Sample_6
Value of the data sample	2	3	5	6	6	8

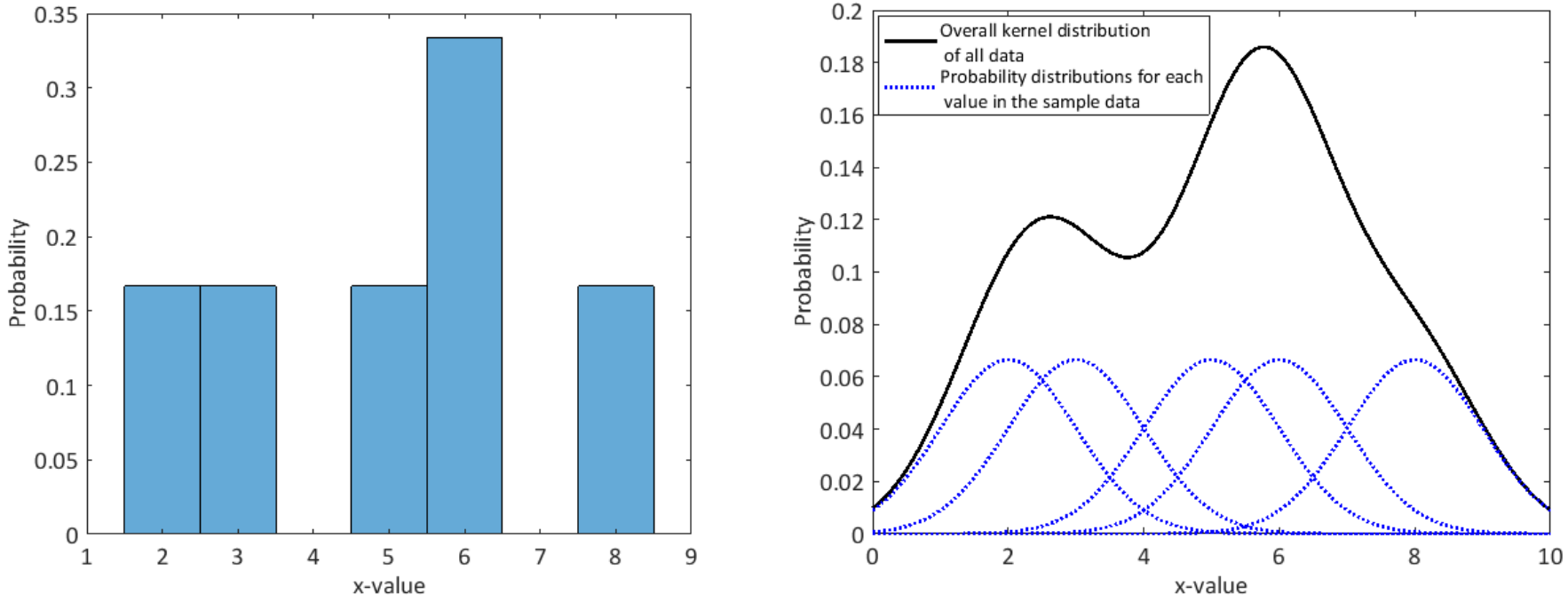


Figure A. 1: (a) Histogram of the selected test data set. (b) The kernel density estimate that was constructed using the same test data set.

For any real values of x, the kernel density estimator's formula is given by,

$$f_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h}) \tag{A.1}$$

$$K(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}} \tag{A.2}$$

where, n is the sample size, $K(.)$ is the kernel smoothing function and h is the bandwidth. The kernel smoothing function defines the shape of the curve used to generate the pdf. For example, Equation (A.2) shows a normal distribution as the kernel smoothing function. A range of kernel smoothing functions are commonly used: uniform, triangular, biweight, triweight, Epanechnikov, normal, and others.

The choice of bandwidth value h, controls how wide the probability mass is spread around a point and hence, controls the smoothness of the resulting probability density curve. Therefore, bandwidth selection bears danger of under or over smoothing of the resulting probability density curve. Figure A.2 shows a comparison of kernel density estimates constructed using the same data set in Table A.1, with different bandwidth selections of the selected (normal) kernel smoothing function. According to the Figure A.2, for a 'normal' kernel smoothing function with $h = 2$, has resulted in an over-smoothed kernel density estimate while $h = 0.15$ has resulted in an under-smoothed kernel density estimate.

There are a great number of bandwidth selection techniques for kernel density estimator. The integrated squared error and the mean integrated squared error have been widely used as the optimality criterion to define the optimal bandwidth [107].

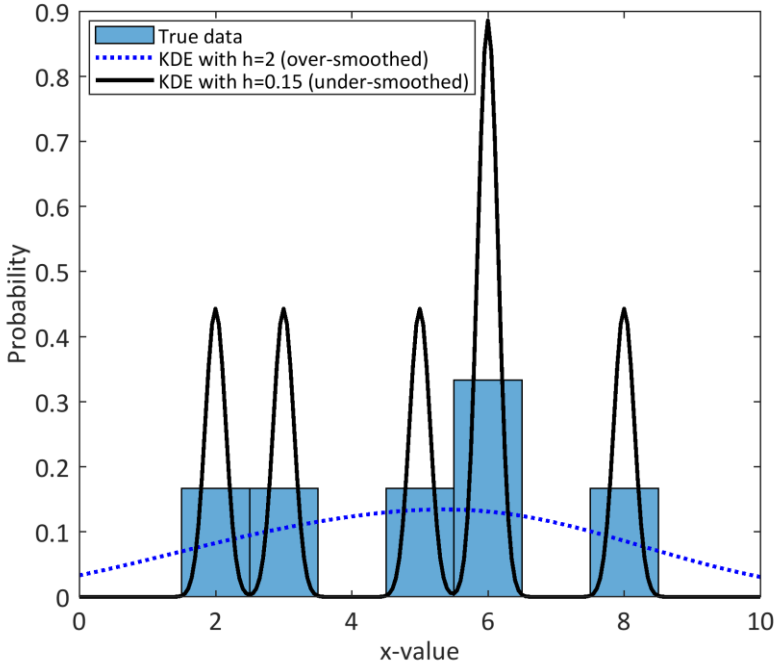


Figure A. 2: Kernel density estimate with different bandwidths.

References

[1] Kyoto Protocol, "United Nations framework convention on climate change," *Kyoto Protocol, Kyoto,* vol. 19, 1997.

[2] European Commission, "A policy framework for climate and energy in the period from 2020 to 2030," *Technical Report COM (2014) 15,* 2014.

[3] G. Pepermans, J. Driesen, D. Haeseldonckx, R. Belmans, and W. D'haeseleer, "Distributed generation: definition, benefits and issues," *Energy Policy,* vol. 33, pp. 787-798, 2005/04/01/ 2005.

[4] T. Chin Ho and G. Chin Kim, "Impact of grid-connected residential PV systems on the malaysia low voltage distribution network," in *Power Engineering and Optimization Conference (PEOCO), 2013 IEEE 7th International*, 2013, pp. 670-675.

[5] L. P. Fernandez, T. G. S. Roman, R. Cossent, C. M. Domingo, and P. Frias, "Assessment of the Impact of Plug-in Electric Vehicles on Distribution Networks," *IEEE Transactions on Power Systems,* vol. 26, pp. 206-213, 2011.

[6] S. Abeysinghe, S. Nistor, J. Wu, and M. Sooriyabandara, "Impact of Electrolysis on the Connection of Distributed Generation," *Energy Procedia,* vol. 75, pp. 1159-1164, 2015/08/01 2015.

[7] F. J. Ruiz-Rodriguez, Herna, x, J. C. ndez, and F. Jurado, "Probabilistic load flow for radial distribution networks with photovoltaic generators," *Renewable Power Generation, IET,* vol. 6, pp. 110-121, 2012.

[8] D. Q. Hung, Z. Y. Dong, and H. Trinh, "Determining the size of PHEV charging stations powered by commercial grid-integrated PV systems considering reactive power support," *Applied Energy,* vol. 183, pp. 160-169, 2016.

[9] A. Rabiee and S. M. Mohseni-Bonab, "Maximizing hosting capacity of renewable energy sources in distribution networks: A multi-objective and scenario-based approach," *Energy,* vol. 120, pp. 417-430, 2017/02/01/ 2017.

[10] K. Clement-Nyns, E. Haesen, and J. Driesen, "The Impact of Charging Plug-In Hybrid Electric Vehicles on a Residential Distribution Grid," *IEEE Transactions on Power Systems,* vol. 25, pp. 371-380, 2010.

[11] C. Gonzalez, J. Geuns, S. Weckx, T. Wijnhoven, P. Vingerhoets, T. D. Rybel*, et al.*, "LV distribution network feeders in Belgium and power quality issues due to increasing PV penetration levels," in *2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, 2012, pp. 1-8.

[12] J. D. Watson, N. R. Watson, D. Santos-Martin, A. R. Wood, S. Lemon, and A. J. Miller, "Impact of solar photovoltaics on the low-voltage distribution network in New Zealand," *IET Generation, Transmission & Distribution,* vol. 10, pp. 1-9, 2016.

[13] J. D. W. N. R. Watson, R. M. Watson, K. Sharma, A. Miller, "Impact of Electric Vehicle Chargers on a Low Voltage Distribution System," presented at the EEA Conference & Exhibition 2015, Wellington, 2015.

[14] P. Mancarella, G. Chin Kim, and G. Strbac, "Evaluation of the impact of electric heat pumps and distributed CHP on LV networks," in *PowerTech, 2011 IEEE Trondheim*, 2011, pp. 1-7.

[15] G. Chin Kim, D. Pudjianto, P. Djapic, and G. Strbac, "Strategic Assessment of Alternative Design Options for Multivoltage-Level Distribution Networks," *Power Systems, IEEE Transactions on,* vol. 29, pp. 1261-1269, 2014.

[16] G. Davies, G. Prpich, N. Strachan, and S. Pollard, "UKERC Energy Strategy Under Uncertainties: Identifying Techniques for Managing Uncertainty in the Energy Sector," UKERC Working Paper UKERC/WP/FG/2014/0012014.

[17] M. Catrinu, "Decision Aid for Planning Local Energy Systems: Application of Multi-Criteria Decision Analysis," 2006.

[18] S. Carr, G. C. Premier, A. J. Guwy, R. M. Dinsdale, and J. Maddy, "Energy storage for active network management on electricity distribution networks with wind power," *IET Renewable Power Generation,* vol. 8, pp. 249-259, 2014.

[19] M. A. Zehir, A. Batman, M. A. Sonmez, A. Font, D. Tsiamitros, D. Stimoniaris*, et al.*, "Impacts of microgrids with renewables on secondary distribution networks," *Applied Energy,* vol. 201, pp. 308-319, 2017/09/01/ 2017.

[20] Z. Wang, A. Scaglione, and R. J. Thomas, "Generating statistically correct random topologies for testing smart grid communication and control networks," *IEEE transactions on Smart Grid,* vol. 1, pp. 28-39, 2010.

[21] G. A. Pagani and M. Aiello, "Towards Decentralization: A Topological Investigation of the Medium and Low Voltage Grids," *Smart Grid, IEEE Transactions on,* vol. 2, pp. 538-547, 2011.

[22] P. Crucitti, V. Latora, and M. Marchiori, "A topological analysis of the Italian electric power grid," *Physica A: Statistical Mechanics and its Applications,* vol. 338, pp. 92-97, 7/1/ 2004.

[23] K. Atkins, J. Chen, V. S. Anil Kumar, and A. Marathe, "The structure of electrical networks: a graph theory based analysis," *International Journal of Critical Infrastructures,* vol. 5, pp. 265-284, 2009/01/01 2009.

[24] R. Albert, I. Albert, and G. L. Nakarado, "Structural vulnerability of the North American power grid," *Physical review E,* vol. 69, p. 025103, 2004.

[25] P. Hines, S. Blumsack, E. Cotilla Sanchez, and C. Barrows, "The Topological and Electrical Structure of Power Grids," in *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, 2010, pp. 1-10.

[26] E. Bompard, D. Wu, and F. Xue, "Structural vulnerability of power systems: A topological approach," *Electric Power Systems Research,* vol. 81, pp. 1334-1340, 7// 2011.

[27] M. ROSAS-CASALS, S. VALVERDE, and R. V. SOLÉ, "TOPOLOGICAL VULNERABILITY OF THE EUROPEAN POWER GRID UNDER ERRORS AND ATTACKS," *International Journal of Bifurcation and Chaos,* vol. 17, pp. 2465-2475, 2007.

[28] D. P. Chassin and C. Posse, "Evaluating North American electric grid reliability using the Barabási–Albert network model," *Physica A: Statistical Mechanics and its Applications,* vol. 355, pp. 667-677, 9/15/ 2005.

[29] V. Rosato, S. Bologna, and F. Tiriticco, "Topological properties of high-voltage electrical transmission networks," *Electric Power Systems Research,* vol. 77, pp. 99-105, 2007/02/01/ 2007.

[30] Y. Koç, M. Warnier, P. Van Mieghem, R. E. Kooij, and F. M. Brazier, "The impact of the topology on cascading failures in a power grid model," *Physica A: Statistical Mechanics and its Applications,* vol. 402, pp. 169-179, 2014.

[31] F. Barakou, D. Koukoula, N. Hatziargyriou, and A. Dimeas, "Fractal geometry for distribution grid topologies," in *IEEE Eindhoven PowerTech, PowerTech 2015*, 2015.

[32] G. Andersson, P. Donalek, R. Farmer, N. Hatziargyriou, I. Kamwa, P. Kundur*, et al.*, "Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance," *IEEE transactions on Power Systems,* vol. 20, pp. 1922-1928, 2005.

[33] M. Van Steen, *Graph Theory and Complex Networks*, 2010.

[34] B. A.-L. BARABÁSI and E. Bonabeau, "Scale-free," *Scientific American,* 2003.

[35] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *nature,* vol. 406, p. 378, 2000.

[36] B. B. Mandelbrot, *The Fractal Geometry of Nature*: 1997, 1982.

[37] J. J. Grainger and W. D. Stevenson, *Power system analysis* vol. 621: McGraw-Hill New York, 1994.

[38] R. Molontay, "Networks and fractals," Department of Stochastics, Budapest University of Technology and Economics, 2013.

[39] G. A. Pagani and M. Aiello, "The Power Grid as a complex network: A survey," *Physica A: Statistical Mechanics and its Applications,* vol. 392, pp. 2688-2700, 6/1/ 2013.

[40] G. A. Pagani and M. Aiello, "Power grid complex network evolutions for the smart grid," *Physica A: Statistical Mechanics and its Applications,* vol. 396, pp. 248-266, 2/15/ 2014.

[41] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, 2005, pp. 177-187.

[42] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani, "Kronecker graphs: An approach to modeling networks," *Journal of Machine Learning Research,* vol. 11, pp. 985-1042, 2010.

[43] D. Chakrabarti, Y. Zhan, and C. Faloutsos, "R-MAT: A recursive model for graph mining," in *Proceedings of the 2004 SIAM International Conference on Data Mining*, 2004, pp. 442-446.

[44] J. E. C. Sánchez, "A complex network approach to analyzing the structure and dynamics of power grids," The University of Vermont, 2009.

[45] Å. J. Holmgren, "Using graph models to analyze the vulnerability of electric power networks," *Risk analysis,* vol. 26, pp. 955-969, 2006.

[46] S. Mei, X. Zhang, and M. Cao, *Power grid complexity*: Springer Science & Business Media, 2011.

[47] K. Sun, "Complex networks theory: A new method of research in power grid," in *Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES*, 2005, pp. 1-6.

[48] J. Ding, X. Bai, W. Zhao, Z. Fang, Z. Li, and M. Liu, "The improvement of the small-world network model and its application research in bulk power system," in *Power System Technology, 2006. PowerCon 2006. International Conference on*, 2006, pp. 1-5.

[49] E. Jelenius, "Graph Models of Infrastructures and the Robustness of Power Grids," *Master of Science in Physics Engineering,* 2004.

[50] G. J. Correa and J. M. Yusta, "Grid vulnerability analysis based on scale-free graphs versus power flow models," *Electric Power Systems Research,* vol. 101, pp. 71-79, 2013.

[51] P. Erdős and A. Rényi, "On the strength of connectedness of a random graph," *Acta Mathematica Hungarica,* vol. 12, pp. 261-267, 1961.

[52] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature,* vol. 393, pp. 440-442, 1998.

[53] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science,* vol. 286, pp. 509-512, 1999.

[54] L. Tao, P. Wenjiang, W. Shaoping, and H. Zhenya, "Topological multifractal in realistic networks with explain by diffusion limited aggregation mechanism," in *Neural Networks and Signal Processing, 2008 International Conference on*, 2008, pp. 539-543.

[55] G. Palla, L. Lovász, and T. Vicsek, "Multifractal network generator," *Proceedings of the National Academy of Sciences,* vol. 107, pp. 7640-7645, 2010.

[56] S. D. Bedrosian, D. L. Jaggard, and X. Sun, "Multi-foci fractal-graph networks," in *Circuits and Systems, 1990., IEEE International Symposium on*, 1990, pp. 2674-2676 vol.4.

[57] S. D. Bedrosian and D. L. Jaggard, "A fractal-graph approach to large networks," *Proceedings of the IEEE,* vol. 75, pp. 966-968, 1987.

[58] D. Melovic and G. Strbac, "Statistical model for design of distribution network," in *Power Tech Conference Proceedings, 2003 IEEE Bologna*, 2003, p. 5 pp. Vol.3.

[59] W. Zhifang, R. J. Thomas, and A. Scaglione, "Generating Random Topology Power Grids," in *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*, 2008, pp. 183-183.

[60] J. Hu, "Cluster-and-Connect: An Algorithmic Approach to Generating Synthetic Electric Power Network Graphs," Arizona State University, 2015.

[61] P. A. Rikvold, I. A. Hamad, B. Israels, and S. V. Poroseva, "Modeling power grids," *Physics Procedia,* vol. 34, pp. 119-123, 2012.

[62] M. E. Newman, "Assortative mixing in networks," *Physical review letters,* vol. 89, p. 208701, 2002.

[63] F. Barakou, D. Koukoula, N. Hatziargyriou, and A. Dimeas, "Fractal geometry for distribution grid topologies," in *PowerTech, 2015 IEEE Eindhoven*, 2015, pp. 1-6.

[64] A. Block, W. Von Bloh, and H. Schellnhuber, "Efficient box-counting determination of generalized fractal dimensions," *Physical Review A,* vol. 42, p. 1869, 1990.

[65] J.-C. Chen, "Dijkstra's shortest path algorithm," *Journal of Formalized Mathematics,* vol. 15, pp. 144-157, 2003.

[66] E. Lakervi and E. J. Holmes, *Electricity distribution network design*: IET, 1995.

[67] United Nations, Demographic Yearbook (Table 6). (2013).

[68] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, Berkeley, Calif., 1967, pp. 281-297.

[69] A. Al-Wakeel, J. Wu, and N. Jenkins, "k-means based load estimation of domestic smart meter measurements," *Applied Energy,* vol. 194, pp. 333-342, 5/15/ 2017.

[70] B. Awerbuch and R. Gallager, "A new distributed algorithm to find breadth first search trees," *IEEE Transactions on Information Theory,* vol. 33, pp. 315-322, 1987.

[71] M. Ponnavaikko and K. P. Rao, "An approach to optimal distribution system planning through conductor gradation," *IEEE Transactions on Power Apparatus and Systems,* pp. 1735-1742, 1982.

[72] M. Kashem, V. Ganapathy, and G. Jasmon, "Network reconfiguration for load balancing in distribution networks," *IEE Proceedings-Generation, Transmission and Distribution,* vol. 146, pp. 563-567, 1999.

[73] H. Chiang and R. Jean-Jumeau, "Optimal network reconfigurations in distribution systems. I. A new formulation and a solution methodology," *IEEE Transactions on Power Delivery,* vol. 5, pp. 1902-1909, 1990.

[74] Zhejiang Chint Electric Cables Co. Ltd, Wire and Cable series. Available: http://en.chintelectric.com/wwwroot/images/upload/private/12/1%20Products%20PDF/Power%20Cables%20&%20Wires.pd.

[75] The guide for planning and design of distribution, DL/T 5729-2016, Chinese Electricity & Power Standards.

[76] A. Z. Zambom and R. Dias, "A review of kernel density estimation with applications to econometrics," December 2012

[77] S. M. Ross, *Introduction to probability and statistics for engineers and scientists*: Academic Press, 2014.

[78] W. J. Stewart, *Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling*: Princeton University Press, 2009.

[79] Weisstein, Eric W. "Independent Statistics." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/IndependentStatistics.html.

[80] M. Ponnavaikko and K. S. P. Rao, "Optimal Distribution System Planning," *IEEE Transactions on Power Apparatus and Systems,* vol. PAS-100, pp. 2969-2977, 1981.

[81] H. N. Tram and D. L. Wall, "Optimal conductor selection in planning radial distribution systems," *IEEE Transactions on Power Systems,* vol. 3, pp. 200-206, 1988.

[82] R. McGill, J. W. Tukey, and W. A. Larsen, "Variations of box plots," *The American Statistician,* vol. 32, pp. 12-16, 1978.

[83] W. Cao, J. Wu, and N. Jenkins, "Feeder load balancing in MV distribution networks using soft normally-open points," in *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2014 IEEE PES*, 2014, pp. 1-6.

[84] A. Kechroud, J. Myrzik, and W. Kling, "Taking the experience from flexible AC transmission systems to flexible AC distribution systems," in *Universities Power Engineering Conference, 2007. UPEC 2007. 42nd International*, 2007, pp. 687-692.

[85] F. Blaabjerg, Z. Chen, and S. B. Kjaer, "Power electronics as efficient interface in dispersed power generation systems," *IEEE transactions on power electronics,* vol. 19, pp. 1184-1194, 2004.

[86] J. M. Bloemink and T. C. Green, "Increasing distributed generation penetration using soft normally-open points," in *IEEE PES General Meeting*, 2010, pp. 1-8.

[87] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Transactions on Power Delivery,* vol. 4, pp. 1401-1407, 1989.

[88] W. Cao, J. Wu, N. Jenkins, C. Wang, and T. Green, "Benefits analysis of Soft Open Points for electrical distribution network operation," *Applied Energy,* vol. 165, pp. 36-47, 2016/03/01/ 2016.

[89] W. Cao, J. Wu, N. Jenkins, C. Wang, and T. Green, "Operating principle of Soft Open Points for electrical distribution network operation," *Applied Energy,* vol. 164, pp. 245-257, 2016/02/15/ 2016.

[90] J. M. Bloemink and T. C. Green, "Increasing photovoltaic penetration with local energy storage and soft normally-open points," in *2011 IEEE Power and Energy Society General Meeting*, 2011, pp. 1-8.

[91] C. Long, J. Wu, L. Thomas, and N. Jenkins, "Optimal operation of soft open points in medium voltage electrical distribution networks with distributed generation," *Applied energy,* vol. 184, pp. 427-437, 2016.

[92] Q. Qi, J. Wu, and C. Long, "Multi-objective operation optimization of an electrical distribution network with soft open point," *Applied Energy,* vol. 208, pp. 734-744, 2017.

[93] J. D. Watson, N. R. Watson, A. Miller, D. Santos-Martin, S. Lemon, and A. Wood, "Low Voltage Network Modelling," 2014.

[94] K. Zou, A. P. Agalgaonkar, K. M. Muttaqi, and S. Perera, "Distribution system planning with incorporating DG reactive capability and system uncertainties," *IEEE Transactions on Sustainable Energy,* vol. 3, pp. 112-123, 2012.

[95] B. Piccoli and F. Rossi, "Generalized Wasserstein distance and its application to transport equations with source," *Archive for Rational Mechanics and Analysis,* vol. 211, pp. 335-358, 2014.

[96] C. Wang, G. Song, P. Li, H. Ji, J. Zhao, and J. Wu, "Optimal siting and sizing of soft open points in active electrical distribution networks," *Applied Energy,* vol. 189, pp. 301-309, 2017/03/01/ 2017.

[97] S. H. Jangamshetti and V. Ran, "Optimum siting of wind turbine generators," *IEEE Transactions on Energy Conversion,* vol. 16, pp. 8-13, 2001.

[98] V. Graham and K. Hollands, "A method to generate synthetic hourly solar radiation globally," *Solar Energy,* vol. 44, pp. 333-341, 1990.

[99] M. Baran and F. F. Wu, "Optimal sizing of capacitors placed on a radial distribution system," *IEEE Transactions on power Delivery,* vol. 4, pp. 735-743, 1989.

[100] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of machine learning*, ed: Springer, 2011, pp. 760-766.

[101] J. MATHEWS and K. Fink, "Module for the Powell Search Method for a Minimum," *California State Univ. Fullerton-CSUF Math. Dept Website,* 2003.

[102] M. Nick, R. Cherkaoui, and M. Paolone, "Optimal allocation of dispersed energy storage systems in active distribution networks for energy balance and grid support," *IEEE Transactions on Power Systems,* vol. 29, pp. 2300-2310, 2014.

[103] T. Ackermann, N. B. Negra, J. Todorovic, and L. Lazaridis, "Evaluation of electrical transmission concepts for large offshore wind farms," in *Copenhagen Offshore Wind Conference and Exhibition, Copenhagen*, 2005, pp. 26-28.

[104] M. A. Parker and O. Anaya-Lara, "Cost and losses associated with offshore wind farm collection networks which centralise the turbine power electronic converters," *IET Renewable Power Generation,* vol. 7, pp. 390-400, 2013.

[105] M. Trutschl, G. Grinstein, and U. Cvek, "Intelligently resolving point occlusion," in *Information Visualization, 2003. INFOVIS 2003. IEEE Symposium on*, 2003, pp. 131-136.

[106] MathWorks. (Retrieved 30th July, 2018). *Statistics and Machine Learning Toolbox: Documentation (R2018a)*. Available: https://uk.mathworks.com/help/stats/kernel-distribution.html

[107] B. A. Turlach, "Bandwidth selection in kernel density estimation: A review," in *CORE and Institut de Statistique*, 1993.